

Open Research Online

The Open University's repository of research publications
and other research outputs

Dinucleotide TG Repeats and 5' Splice Site Definition

Thesis

How to cite:

Passoni, Monica (2011). Dinucleotide TG Repeats and 5' Splice Site Definition. PhD thesis The Open University.

For guidance on citations see [FAQs](#).

© 2011 The Author

Version: Version of Record

Copyright and Moral Rights for the articles on this site are retained by the individual authors and/or other copyright owners. For more information on Open Research Online's data [policy](#) on reuse of materials please consult the policies page.

oro.open.ac.uk

Dinucleotide TG repeats and 5' splice site definition

Monica Passoni

**A Thesis Submitted in Fulfillment of the Requirements for the Degree
of Doctor of Philosophy in Life Sciences**

**ICGEB, Italy
The Open University, UK**

**Director of Studies: Prof. Francisco E. Baralle, M.D. Ph.D.
External Supervisor: Prof. Robin Reed, Ph.D.**

June 2011

DATE OF SUBMISSION: 18 MAY 2011

DATE OF AWARD: 2 SEPT 2011

ProQuest Number: 13837559

All rights reserved

INFORMATION TO ALL USERS

The quality of this reproduction is dependent upon the quality of the copy submitted.

In the unlikely event that the author did not send a complete manuscript and there are missing pages, these will be noted. Also, if material had to be removed, a note will indicate the deletion.



ProQuest 13837559

Published by ProQuest LLC (2019). Copyright of the Dissertation is held by the Author.

All rights reserved.

This work is protected against unauthorized copying under Title 17, United States Code
Microform Edition © ProQuest LLC.

ProQuest LLC.
789 East Eisenhower Parkway
P.O. Box 1346
Ann Arbor, MI 48106 – 1346

Abstract

During the pre-mRNA splicing process introns are removed and exons joined together in the resulting mature mRNA, which is then exported to the cytoplasm and translated into proteins. Several motifs in the nucleotide sequences near the exon-intron boundaries are required for proper exon definition including the 3' and 5' splice site consensus sequences (3'SS and 5'SS), which recruit basic splicing factors. In addition, auxiliary splicing regulatory elements, located in the upstream or downstream region of an exon, further influence exon recognition through the recruitment of additional binding proteins.

This study shows that intronic UG repeat elements in proximity of the 5'SS of an exon can function as splicing regulatory elements, generally enhancing the inclusion of the upstream exon in the final mRNA through the recruitment of UG repeats-binding proteins. In particular, the strength of the 5'SS consensus sequence affects this UG repeats-mediated splicing regulation. Furthermore, this study reveals that the UG repeats-binding protein TDP-43 acts as splicing modulator either activating or inhibiting the splicing events in different minigene systems. In fact, in presence of a disease-causing mutation at the 5' end of the BRCA1 exon 12 TDP-43 enhances exon inclusion, acting as splicing enhancer. Alternatively, overexpression of TDP-43 can exert an inhibitory effect on splicing by promoting exon skipping in two newly identified TDP-43 target exons (RXRG exon 7 and ETF1 exon 7).

In conclusion, this study provides positive proof of concept that UG repeats located in the downstream region of poorly defined exons help 5'SS definition. Additionally, the study characterizes the role of TDP-43 in various splicing systems presenting the UG repeat-binding sites, with the practical application of evaluating a putative splicing-affecting pathological mutation.

Table of Contents

Abstract	ii
Table of Contents	iii
Introduction	1
Elements required for exon recognition	1
Repeats, repeats, repeats	4
Tandem repeats	5
Characteristics	5
Localization	7
Positive effects: TR relevance in evolution	8
Negative effects: TRs relevance in human pathologies..	9
TRs-mediated splicing regulation	10
Dinucleotide repeats	12
Global distribution, classification and role	12
Dinucleotide repeats form Z-DNA and Z-RNA	13
TG repeat polymorphisms	15
TG repeats in gene expression	20
UG repeats-mediated splicing regulation	24
TDP-43	30
The protein	30
Localization	32
Mis-localization and TDP-43 Proteinopathies	33
Binding specificity	35
RNA targets overview	38
RNA-binding partners	40
Overview of TDP-43 nuclear functions	45
TDP-43 in transcription	45
TDP-43 in general gene expression	47
TDP-43 in splicing regulation	49
TDP-43 in microRNAs biogenesis	57
Overview of TDP-43 cytoplasmatic functions	59
TDP-43 in RNA transport granules	59
TDP-43 in P-bodies	61
TDP-43 in stress granules	62

Aim of the project	63
Materials and Methods	64
Constructs preparation	64
<i>In vitro</i> traditional splicing assay	73
<i>In vitro</i> coupled RNAP II transcription and splicing	74
<i>In vitro</i> splicing assay in cell culture	74
cDNA synthesis and analysis	75
Western blots and antibodies	76
Spliceosomal assembly analysis	77
<i>In silico</i> analysis and predictions	77
CLIP analysis	78
Statistical Analysis	79
Results.....	81
Influence of UG repeats length and distance from the 5'SS on exon processing	81
Correlation between UG repeats effect and 5'SS strength	89
Effect of UG repeats-binding proteins on BRCA1 exon 12	97
Effect of TDP-43 levels on BRCA1 exon 12 splicing regulation	106
Identification of new TDP-43 potential RNA targets	117
TDP-43 in splicing regulation of five new potential targets	122
Potential TDP-43 involvement in endogenous 5'SS definition	138
Discussion and Future Directions.....	142
<i>In vitro</i> minigene-based RNA analysis	142
Endogenous RNA analysis	152
References	154
Appendices	170
Oligonucleotides List	170
Figures and Tables List	173
Supplementary Figures	175

Introduction

Elements required for exon recognition

Expression of protein-coding genes in mammals is a multi-step process, beginning with transcription of pre-mRNAs by the RNAPII enzyme (RNA polymerase II) in the nucleus. The pre-mRNAs undergo several processing steps, including splicing, that are physically and functionally coupled to one another [1-5]. During pre-mRNA splicing introns are removed and exons joined together in the resulting mature mRNA. Several motifs in the nucleotide sequences near the exon/intron boundaries are required for proper exon definition, including 3' and 5' splice site consensus sequences (3'SS and 5'SS) and auxiliary splicing regulatory elements.

The 5' splice site consensus sequences

Mammalian 5'SSs comprises nine partially conserved nucleotides at the exon/intron joint: MAG/GURAGU (M is Adenine or Cytosine and R is Adenine or Guanine), which nearly correspond to a Watson and Crick base pairing to the U1 snRNA 5' end (Figure 1). For each 5'SS, the homology to U1snRNA is used to defined the 5'SS strength, which can be measured with different bioinformatic tools in terms of consensus values (CV). Optimal 5'SSs presenting perfect homology to U1 snRNA (CV=1=100%) are rare in human genes [6] and the average CV of authentic splice sites generally spans from 77 to 89% (0.77-0.89) [7].

The 3' splice site consensus sequences

The 3'SSs are defined by three separate sequences: the branch site, the

polypyrimidine tract and the nearly invariant AG dinucleotide at the 3' end of the intron (Figure 1). The branch site comprises seven poorly conserved nucleotides: YNYURAC (where Y is a pyrimidine and N is any nucleotide), which recruit U2 snRNA. The invariant adenosine (A) residue within the sequence, which is usually located 18–40 bases from the 3' end [8, 9], represents the branch point forming the RNA branch in the lariat structure. The pyrimidine-rich region helps the recognition of the 3' end through the recruitment of U2AF (U2 auxiliary factors). In general, the extent of the polypyrimidine tract defines the strength of the 3'SSs: long polypyrimidine tracts insure high affinity binding sites for spliceosomal components and promote efficient exon recognition [10]. Generally, the first AG dinucleotide downstream to the branch site and the polypyrimidine tract are the most commonly used as 3' end.

Exonic/intronic splicing enhancers/silencers

Auxiliary splicing regulatory elements, known as exonic or intronic splicing enhancers/silencers (ESE/ISE and ESS/ISS), further influence exons recognition. These *cis*-acting regulatory sequences generally recruit auxiliary proteins involved in splicing regulation referred to as *trans*-acting factors (i.e. SR proteins - serine/arginine-rich proteins; hnRNP proteins - heterogeneous nuclear ribonucleoproteins; other splicing factors). The combinatory effect of all the *cis*- and *trans*-elements located in proximity or within an exon affects the spliceosome assembly, ultimately promoting or inhibiting exon inclusion in the final mRNA (reviewed in [11]).

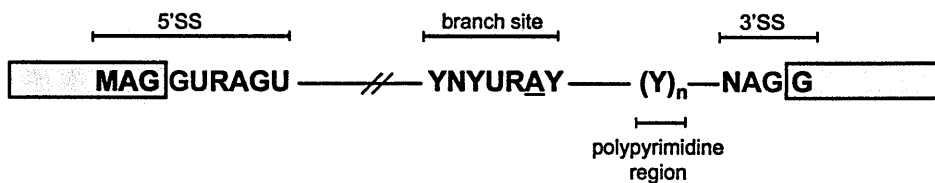


Figure 1: Elements required for proper exon definition. The 3' splice site consensus sequence (defined by the branch site, the polypyrimidine region and the nearly invariant AG dinucleotide) together with the 5' splice site consensus sequence ensures proper exon recognition. In addition, auxiliary splicing regulatory elements (not illustrated in this figure) further help exon definition.

This dissertation focuses on a potential *cis*-regulatory element, the dinucleotides TG repeats, and a specific *trans*-acting factor, TDP-43 (TAR DNA-binding protein 43kDa).

Repeats, repeats, repeats

A puzzling observation in the early days of molecular biology was that genome size does not always correlate well with organism complexity. For example, human genome is 200 times larger than yeast genome, but 200 times as small as that of amoeba. This mystery has been largely solved with the recognition that genomes can contain a large quantity of repetitive sequences, far in excess of that devoted to protein-coding genes [12, 13]. Although extremely abundant in prokaryotes and eukaryotes, repeats were initially believed to have no biological function and therefore labeled as junk or selfish DNA [14, 15]. Conversely, it is now recognized that they actually represent an extraordinary trove of information about biological processes. In fact, repeats are markers of crucial evolutionary events and provide a tool for medical and population genetic studies. Additionally, certain repeats have been linked with several aspects of gene expression such as transcription, mRNA processing and translation.

Repeated DNA sequences account for at least 50 % of the human genome and broadly fall into five classes: 1) transposable elements, often referred to as interspersed repeats; 2) retropseudogenes; 3) segmental duplication; 4) blocks of repeated sequences, such as at centromeres and telomeres and 5) tandem repeats [13, 16]. Among them, the class of tandem repeats and its subclass of dinucleotide TG repeats in particular is thoroughly reviewed in the next pages.

Tandem repeats

Characteristics

Tandem repeats (TRs) are short DNA repeated sequences located right next to the other (in tandem), generally thought to arise by slippage during DNA replication. Although some are conserved over considerable evolutionary distances implying an advantageous effect of some kind, TRs are extremely unstable [17].

Their classification is based on characteristic such as unit length, repeat unit number and repeat purity (Figure 2).

The “unit length” represents the number of repeated nucleotides within a unit. When the unit length ranges from 1 to 9 nucleotides, the TRs are called microsatellites (or short tandem repeats). Among them are for example mononucleotide repeats (unit length equal to 1), dinucleotide repeats (unit length equal to 2) and trinucleotide repeats (unit length equal to 3). Alternatively, when the unit length is greater than 10 nucleotides TRs are defined as minisatellites.

The “repeat unit number” (or repeat length) represents the number of times a single unit is repeated within a TR and it is also the most important factor defining TRs instability. In general, the higher the repeat unit number, the more unstable the TR. The variability in repeat unit number is defined as length polymorphism, a feature largely used for human genetic studies. In fact, addition or deletion of entire units is a frequent event that provides useful markers and is also capable of inducing different phenotypes.

Finally, the “repeat purity” discerns between perfect repeats, where a unit is uninterruptedly repeated, and imperfect repeats, which instead have accumulated mutations within their units (Figure 2).

Unit Length

G G G G
unit

unit length = 1 nt

TG TG TG TG
unit

unit length = 2 nt

CTG CTG CTG CTG
unit

unit length = 3 nt

Microsatellites: unit length = 1 to 9 nt

Example: CCTG CCTG CCTG CCTG CCTG

Minisatellites: unit length greater than 10 nt

Example: GTACACTTTACC GTACACTTTACC GTACACTTTACC

Repeat Length

TG TG TG TG TG TG

repeat length = 6

CTG CTG CTG

repeat length = 3

Repeat Purity

Perfect repeat

Example: TGTGTGTGTGTGTGTGTGTG
(NTRK1 intron 4)

Purity

100%

Imperfect repeat

Example 1: TGTGTGTCTGTGTGTATGTG
(MYL intron 5)

90%

Example 2: TGTGTTGGGTGTTGTGTGTG
(EPHX1 intron 6)

74%

Figure 2: Tandem repeats characteristics. TRs are classified based on their unit length, repeat unit number and repeat purity.

Interestingly, TRs comprise about 3% of the human genome and are almost as abundant as coding sequences, which account for less than 5% of it [13, 18]. The greatest single contribution is coming from the dinucleotide repeats (0.5%), followed by tetranucleotides (unit length equal to 4) and pentanucleotides (unit length equal to 5).

Localization

TRs have been shown to localize both in coding and non-coding regions where they exert disparate biological functions. Polymorphisms of the repeated regions have great impact on their functionality in both locations.

The presence of TRs in the coding sequence results in a stretch of amino acid residues (i.e. trinucleotide CAG repeats encode for polyglutamine stretch; dinucleotide TG repeats encode for alternating cysteine and valine stretches). Within the coding sequence, repeat length variations result in the production of an altered protein. This event can either functionally improve the protein and positively affect phenotypes, or alternatively result in a non-functional protein triggering disease pathogenesis.

On the other hand, TRs localized in non-coding regulatory elements such as promoter regions, 5'/3'UTRs and introns, modulate gene expression by functioning as protein binding sites or by affecting DNA and RNA secondary structures. Consequently, in a variety of different ways, TRs are implicated in transcription activation, mRNA stability, constitutive and alternative splicing, translation and represent also signals for mRNA export [17, 19-23]. Consequently, alteration of the repeat unit number can interfere with all these TRs functions by:

- i) disrupting or strengthening binding sites for regulatory proteins [24-26];
- ii) changing the spacing between different regulatory sites [27];

iii) influencing chromatin structure and therefore the accessibility of transcription factors and regulatory proteins ultimately interfering with nucleosome formation [28];

iv) affecting DNA secondary structure leading to the formation of Z-DNA [29], trapping RNAPII on the transcript and influencing transcription initiation and elongation [30];

v) affecting RNA secondary structure leading to the formation of Z-RNA or extremely stable hairpins that influence both pre-mRNA processing and translation [31, 32].

Therefore, as for TRs in coding sequences, perturbation of the TRs length in non-coding regions can have an ultimate beneficial effect or alternatively result in a defective phenotype, or can also be neutral.

Positive effects: TRs relevance in evolution and environmental adaptation

Several examples of how TRs instability can advantageously affect phenotypes are reported in literature and also reviewed in reference number 17 [17]. In two prokaryotic pathogenic bacteria causing gonorrhea and influenza respectively, variations of TRs located in coding regions lead to the production of differential proteins. Alternatively, variations of TRs located in the promoter region modulate gene expression. In this way TRs rapidly mediate the switch between different phenotypes, essential for the organism to evade the host defense system [33, 34].

Also, in eukaryotic microbes TRs located within coding regions are responsible for environmental adaptation. For example, TRs length variability within the adhesin gene was proved to modulate cell adherence to the substrate or the host tissue, both in benign yeast *Saccharomyces cerevisiae* and in the pathogenic fungus *Candida albicans* respectively [35, 36]. Furthermore, expanded

TRs leading to longer polyglutamine (poly-Q) stretches were found to correlate with circadian clocks control in fungi, fly and birds [37-40].

In higher vertebrates, unstable TRs confer evolutionary flexibility to body shape and organs [41] as proved for example in *Canis familiaris* (dog). TRs length polymorphism, in fact, was found to correlate with the midface length and dorsoventral nose bend in different dog breeds [42]. Interestingly, mutations in the TRs of the corresponding human gene were reported to result in craniofacial malformations [17].

Negative effect: TRs relevance in human pathologies

In some contexts, TRs high instability can ultimately result in a defective phenotype. Among the class of TRs, expansions observed in the subclass of trinucleotide repeats are by far the most common cause of human diseases [17]. In particular, expanded trinucleotide CAG repeats located within coding regions cause poly-Q diseases. These neurodegenerative disorders include Huntington's disease (HD), spinobulbar muscular atrophy (SBMA) [43] and spinocerebellar ataxias (SCAs: from SCA1 to SCA17) [44]. The triplet codon CAG corresponds to the amino acid glutamine (Q). Therefore, CAG expansions result in altered protein products carrying long poly-Q tracts. The poly-Q stretches confer self-aggregating properties to the protein ultimately leading to the formation of toxic aggregates in patient's cell.

Trinucleotide expansions in non-coding regions, such as 5'/3'UTRs and introns, are responsible for fragile X syndrome, Friedreich ataxia, and myotonic dystrophy type 1 and 2 (DM1 and 2). In fragile X syndrome and Friedreich ataxia, the most common forms of mental retardation and of ataxia respectively, these expansions affect transcription factor binding and chromatin structure, resulting in transcriptional silencing of the genes involved. Alternatively, in DM1 and DM2

expansions affect the binding of two RNA processing protein (CELF1 and MNBL), resulting in altered splicing of different RNAs [21, 45, 46].

Expansions of polymorphic TRs dinucleotide repeats located in non-coding regions have also been linked to several diseases. Interestingly, dinucleotide repeats within coding sequences are rare, possibly because they are selected against during evolution [17]. Examples of dinucleotide repeats involvement in disease onset include: the CA repeats in non-coding region of the interferon-gamma gene, in which a polymorphism has been associated with rheumatoid arthritis [47]; and the CT repeats in exon 1 of the Norrie-disease gene, which has been associated with retinopathy [48]. Most importantly for this study, numerous examples of the TG repeats involvement in various diseases were identified by thoroughly screening the literature. All these examples are found summarized in Table 1 and also described in the next pages. However, although a great number of studies associate different dinucleotide repeat polymorphisms with disparate diseases, their direct involvement in triggering pathogenesis has not always been fully uncovered as for the trinucleotide repeats.

TRs-mediated splicing regulation

As previously mentioned, TRs can significantly affect splicing efficiency and splice site choice when located in proximity of splice signals. In particular, various TRs (mono-, di-, tri- etc.) affect splicing events by acting as protein-binding sites [49].

For example, mononucleotide G repeats regulates 5' and 3'SSs choice in numerous genes through the recruitment of hnRNP H1 [50-53]. Moreover, G-rich elements were recently found to redundantly protect 5'SS weakened by single nucleotide polymorphisms [54].

Also, dinucleotide CA repeats located at the 5'SS generally function as strong intronic splicing enhancers (ISE) in: eNOS exon 13 [19, 49, 55]; MAPK10 exon 6; GSTZ1 exon 5; SLC2A2 exon 4; RFXANK exon 5 [49]; and CD45 exon 4 [56]. The CA repeats were found to determine either constitutive or alternative splicing efficiency in a length-dependent manner. They specifically recruit the *trans*-acting factor hnRNP L, which ultimately affects mRNA stability and enhances splicing. The CA repeats recruiting hnRNP L were also found to act as splicing silencers when located at the 3'SS of TJPI exon 20, repressing exon inclusion [57]. In addition, when located at gene promoter regions or in the 5' and 3'UTRs the dinucleotide CA repeats affect both mRNA stability and gene expression. Examples include: bcl-2 3'UTR [58]; IGF1 promoter region [59]; cPLA2 promoter region [60]; HMGA2 promoter region [61]; Cyr61 promoter region [62]; and EGFR intron 1 [63]. Further representative examples of TRs-mediated splicing regulation include the followings.

Trinucleotide CUG repeats located at the 3'SS regulates the aberrant alternative splicing of different pre-mRNAs by affecting the binding of CELF1 and MBNL proteins [64, 65].

A tetranucleotide repeats regulates alternative splicing of the alpha2-glycine-receptor and many others neurotransmitter receptor subunits through the recruitment of Nova-1 [66-68].

A pentanucleotide repeat acts as splicing enhancer of the alternatively spliced exon 4 in the calcitonin gene [69].

A hexanucleotide repeat (unit length equal to 6) regulates tissue-specific alternative splicing of the fibronectin exon EIIIB [70].

Finally, a six 13-nucleotide repeats (unit length equal to 13) modulate *Drosophila* doublesex exon 4 splicing through the binding of Tra, Tra2 and 9G8 proteins [71].

Dinucleotide TG repeats

Global distribution, classification and roles

As previously mentioned, dinucleotide repeats are the most common TRs and account for the 0.5% of the entire human genome. Among the dinucleotide repeats, TG repeats are the most frequent (50%), followed by AT (35%), GA (15%) and the under-represented CG (0.1%) [13]. It has to be noted that when screening the genome there are only four possible types of dinucleotide repeats. In fact TG=GT=CA=AC, GA=AG=TC=CT, AT=TA, and CG=GC [18].

The TG or CA (TG/CA) repeats are predominantly located in intronic regions rather than in exons and are found particularly enriched in some gene families. Analysis of six functional classes revealed a correlation between the gene function and the TG/CA global distribution [20, 72, 73]. Genes carrying in their sequence at least one TG/CA repeat with minimum repeat unit number (or repeat length) of six account for: 61% of the cell-communication and signaling pathway genes (including receptors, hormones and growth factors and channel proteins); 52% of the metabolism genes (including proteasome subunits genes and cytochrome p450 superfamily); 52% of the structure and motility genes (including collagen gene family); 47% of the cell cycle genes (including cell cycle genes, chromosomal structure and DNA repair genes); 45% of the information genes (including translation factors, transcription factors, RNA processing and repair); and 30% of the immune response genes (including immunoglobulins and homeostasis maintainers).

Interestingly, the collagen gene family, belonging to the class of structure and motility, has the higher percentage of genes carrying TG/CA repeats (85.6%) [73].

Based on repeat length and biological properties, TG/CA repeats can be categorized in three types: i) type I include short repeats ($n = 1$ to 11) with low propensity to polymorphism; ii) type II include medium length repeats ($n = 12$ to 22), which are likely polymorphic and often involved in transcription regulation; and iii) type III include rare, long ($n > 22$) and highly polymorphic repeats [73].

Dinucleotide UG repeats (in the RNA context TG repeats are referred to as UG repeats), just like other TRs, can influence splicing efficiency and splice site choice when located in proximity of splicing regulatory sequences (5'/3'SSs). However, TG repeats are most frequently localized at gene promoter regions or in 5'UTRs where they are capable of modulating transcription activation (Table 1). In this case, the adoption of unconventional DNA and RNA structures is the most accredited mechanism responsible for dinucleotide repeats-mediated transcription activation [74]. Interestingly, for each location (5'/3'SSs, introns, 5'/3'UTR, gene promoter) the variability of the TG/UG repeats length was found to differentially affect their biological function (Table 1).

Dinucleotide repeats form Z-DNA and Z-RNA

DNA is normally found in a right-handed helix with 10.4 base pair per turn defined as B-DNA (B-form DNA). However, DNA can translate form B to Z-DNA (zigzag DNA), a left-handed helix with 12 base pairs per turn.

This alternative, biologically functional and higher energy form is preferentially adopted by sequences containing alternating purine-pyrimidine dinucleotide repeats (CG, CA, TG) [75, 76]. Z-DNA conformations at the promoter region or near the transcription start site can affect transcription either enhancing or repressing it.

In fact, chromatin-remodeling proteins unwrap DNA from the nucleosomes leaving the DNA negatively supercoiled in Z-conformation. Since nucleosomes

cannot reform on Z-DNA, the site is kept open allowing transcription factors to accumulate and RNAPII transcription initiation to begin [77, 78].

Alternatively, Z-DNA can repress transcription by affecting RNPII movements and transcription factor binding. Retarded transcription leads to a reduced amount of generated transcripts. Microarray and reporter minigene analysis have shown that, in most cases, short TG/CA repeats (repeat length = 12) at the gene promoter downregulate transcription (i.e. TG/CA repeats in HSD11B2 gene) [79]. Consistently, the longer is the repeated element at the promoter region the lower is the production of transcript (i.e. CA repeats in prolactin gene) [20, 73, 80].

It should be noted that Z-DNA, besides acting as *cis*-elements, can also recruit *trans*-acting factors. In fact, a class of Z-DNA binding proteins was identified. Among these proteins, the better characterized is the RNA-editing enzyme ADAR1 [78, 81]. ADAR1 (adenosine deaminase, RNA-specific 1) binds and stabilizes the Z-DNA preventing its return to the B conformation and therefore enhancing transcription [82].

Similarly to DNA, also double stranded RNA can shift from A-RNA to Z-RNA. For example, CG repeats adopt Z-RNA conformation, which can also be bound by ADAR1 protein [83]. Single stranded UG repeats as well adopt Z-RNA conformation, which are bound by CELF1 protein (CUGBP-Elav-like family member 1). Specifically it has been suggested that only the left-handed UG Z-RNA conformation, but not its right-handed counterpart, can be actually accommodated in CELF1 RRM1 and RRM2 pockets (RNA recognition motif 1 and 2). This interaction between CELF1 and the UG in Z-RNA conformation may have possible implications in regulating alternative splicing, mRNA stability and translation [84].

Finally, Z-RNA conformation can also affect translation. In fact, it has been recently reported that rRNAs form Z-RNA structures involved in target recognition. In this case, ADAR1 binds the rRNA Z-structure inhibiting ribosomal complexes formation and thus translation [85].

TG repeats polymorphisms

As previously reported, trinucleotide repeats expansions are involved in the pathogenesis of various neurological diseases (page 9). In order to uncover a similar role for dinucleotide repeats, several genotyping studies have investigated the potential link between polymorphic TG repeats and the susceptibility to certain pathologies (genetic risk). The numerous studies identified in the literature are summarized in Table 1. Indications about the gene involved, the location of the polymorphic TG element and the correlated disease are included. The specific ethnic group where the polymorphism was validated is also reported. Consistent with their predominant distribution, TG repeats analyzed in these studies are mostly localized in non-coding regulatory region. All the studies presented in Table 1 used conventional genotyping methods such as reverse-transcription associated with real-time-PCR to evaluate mRNA levels in RNA extracted from blood or tissue biopsies. In addition, protein expression levels were also evaluated using immunohistochemistry and western blot analysis.

In some notable examples, studies have gone beyond correlating a particular TG allele with a phenotype and actually linked polymorphic TG repeats with transcription or splicing modulation (in bold in Table 1 and summarized in the following paragraphs).

Table 1. TG repeats polymorphisms and pathologies. Summary of the numerous studies investigating the potential link between polymorphic TG repeats and the susceptibility to certain pathologies. The asterisk “*” indicates the genes recently found to be up/downregulated upon TDP-43 silencing (a TG/UG repeats binding protein) in reference 86 [86].

Gene Symbol	Protein Name	TG repeats polymorphism correlations	Location	Ref.
AGT	Angiotensinogen	Correlation with stroke in pediatric patients with sickle cell disease (SCD) in African-American and in African-Caribbean populations.	intron 4, up stream region	[87]
APO	Apolipoprotein A-II	No correlation with plasma concentrations of high-density lipoproteins (HDLs) and cholesterol levels in Irish and French populations. UG repeats regulate splicing.	intron 2, up stream region	[88, 89]
BDNF	Brain-derived neurotrophic factor	Correlation with responsiveness to an antipsychotic agent in schizophrenic patients in Chinese population (pharmaco-genetic relationship).	5'UTR	[90]
		Correlation with schizophrenia in Italian, Canadian and French populations, but not in Japanese, Irish and Spanish populations.		[91-95]
		No association with bipolar disorder susceptibility.		[96]
Beta-ENaC	Epithelial sodium channel beta subunit	Correlation with hypertension in Chilean population. TG repeats regulate gene expression.	intron 8	[97]
BRCA1/2 and RAD51	Breast cancer 1/2 and RAD51	No correlation with breast cancer susceptibility in Polish population.	whole genes	[98]
BRN3C	Brain-specific homeobox/POU domain protein 3C	TG repeats regulate gene expression.	intron 1	[99]
CD36	CD36 molecule thrombospondin receptor	Correlation with cerebral malaria in Thai population. UG repeats regulate alternative splicing.	intron 3, down stream region	[100, 101]
		Correlation with cardiovascular risk and higher body mass index in Korean, American, African-American and British-Caucasian populations.		[102-106]
CDH13	H-cadherin (heart)	Correlation with metastatic cancer: loss of TG repeats detected in lung and breast cancer.	intron 1	[107, 108]

CFTR	Cystic fibrosis transmembrane regulator	Correlation with monosymptomatic cystic fibrosis (CF) in Iranian and Caucasian populations, where allele TG12T7 is the most common. UG repeats modulate alternative splicing. Possible correlation with lower incidence of CF or CF-like diseases in Asian/Chinese population; TG11T7 is the most common allele in these populations.	intron 8, up stream region	[109-111] [112]
CMA1	Mast cell chymase 1	Correlation with atopic asthma and IgE levels.	3'UTR	[113]
D155976	D155976	No association with schizophrenia in Chinese population.	3'UTR	[114]
ER alpha	Estrogen receptor alpha	Correlation with genetic susceptibility to breast cancer. Longer TG tracts associate with reduced expression of progesterone receptor and reduced cancer risk in Chinese and Caucasian populations.	5'UTR	[115-117]
FGF9	Fibroblast growth factor 9	Correlation with gonadal dysgenesis (XY female sex reversal). TG repeats modulate promoter activity, mRNA stability and translation.	3'UTR	[22, 118]
Flt-1	FMS-like tyrosine kinase 1	No association with development of preeclampsia in Korean population.	3'UTR	[119]
FOXP3	forkhead box P3/Scurfin	Correlation with type I diabetes in Japanese population. TG repeats modulate promoter activity. No correlation with several autoimmune diseases (lupus, rheumatoid arthritis, celiac disease, Crohn's disease, etc.) in Spanish population.	5'UTR	[120-122] [123]
GRIN2A *	Glutamate receptor, ionotropic, N-methyl D-aspartate 2A	Correlation with schizophrenia in Chinese population.	promoter region	[124]
H2	Human H2 relaxin gene	Possible correlation with transcription regulation.	5'UTR	[125]
HA	Factor 8 for hemophilia A	Correlation with hemophilia A in Spanish population.	intron 1	[126, 127]
hDBH	Dopamine beta-hydroxylase	No correlation with schizophrenia in Spanish population.		[128]

hERG/ KCNH2	Human ether-a-go-go related gene	Correlation with drug-induced LQTS in Chinese and Caucasian populations. TG repeat affect mRNA and protein levels resulting in QT interval variations.	promoter region	[129]
hGHR	Growth hormone receptor	Correlation with growth disorders in Canadian population.	5'UTR	[130]
HIF1 alpha	Hypoxia-inducible transcription factor 1 alpha	Correlation with non-small cell lung cancer patients and increased HIF1 alpha expression.	intron 13	[131]
HMOX1 *	Eme oxygenase 1	Correlation with coronary artery disease in diabetics and with atherosclerosis after arsenic exposure in Taiwanese population.	promoter region	[132, 133]
		Correlation with hyperbilirubinemia.		[134]
		Correlation with lung adenocarcinoma, oral squamous cancer and gastric adenocarcinoma.		[135]
		No correlation with malaria pathogenicity and severity in general Asiatic populations.		[136]
		Correlation with cerebral ischemic stroke.		[137]
IFNAR1	Interferon alpha receptor 1	Correlation with interferon-induced depressive symptoms.	promoter region	[138]
IL12a/b	Interleukin 12a/b	Correlation with gastric cancer in <i>H. pylori</i> infected individuals in Italian population.	promoter region	[139]
MAOB	Monoamine oxidase A	Correlation with Parkinson's disease in Australian, but not in Chinese population.	intron 2	[140, 141]
		No correlation with bipolar disorder susceptibility in Caucasian population.		[142]
NEP	Neprylisin	No correlation with sporadic Alzheimer's disease susceptibility in Japanese population.	intron 22, exon 1	[143]
		Correlation with sporadic Alzheimer's disease when coexisting with mutations in the angiotensin-converting enzyme in Korean population.		[144]
NOS1	Nitric oxide synthase 1 (neuronal)	Correlation with amount of exhaled nitric oxide concentrations. TG repeats possibly affect gene expression, transcription efficiency and alternative splicing events.	5'UTR	[145]
NTRK1/ TRKa	Neurotrophic tyrosine kinase, receptor, type 1	No correlation with congenital insensitivity to pain with anhidrosis (CIPA). Other polymorphisms in this gene are the direct cause of CIPA disease.	intron 12, down stream region	[146, 147]

p55 *	TNF receptor 1	Correlation with autoinflammatory syndromes, such as familial Iberian fever, in Spanish population.	intron 9 and intron 25	[148, 149]
PHKA1	Phosphorylase kinase, alpha 1; muscle specific	Correlation with phosphorilase kinase deficiency.		[150]
PSMA6	Human proteosome core particle	Correlation with type II diabetes mellitus in Finnish and Latvian population.	intron 6	[151-153]
SCN4A	Sodium channel, voltage-gated, type IV, alpha subunit	Correlation with paramyotonia congenita and hyperkalemic periodic paralysis.		[154]
SLC11A1 *	Solute carrier family 11A member 1	Correlation with autoimmune disease and infection susceptibility. TG repeats modulate gene expression.	promoter region	[155, 156]
		Correlation with tuberculosis development.		[157]
		Correlation with Crohn's disease and inflammatory bowel disease.		[158]
SLC12A1	Solute carrier family 12	No correlation with essential hypertension susceptibility in Kazakhs population.	Not reported	[159, 160]
SLC8A1 *	Solute carrier family 8 member 1	UG repeats modulate alternative splicing.	5'UTR	[161]
STAT6	Signal transducer and activator of transcription 6	Correlation with atopic asthma in British population. TG repeats modulate promoter activity.	5'UTR	[162-164]
Tau	Microtubule-associated protein tau	Correlation with progressive supranuclear palsy (movement disorder).	intron 2, down stream region	[165]
TLR2 *	Toll-like receptor 2	No correlation with bacterial infection susceptibility in Thai population.	5'UTR	[166]
		Correlation with colorectal cancer in Balkan population.		[167]
TNC *	Tenascin-C gene	Correlation with symptomatic Achilles tendon injury.	intron 17, down stream region	[168, 169]
VWF	Von Willebrand factor	TG repeats modulate gene expression in response to shear stress.	5'UTR	[169]

TG repeats in gene expression

As shown in Table 1, according with their location, polymorphic TG dinucleotide repeats are capable of modulating different aspects of gene expression including transcription rate at the promoter, exon recognition efficiency and also mRNA stability. This TG repeats-mediated regulatory effect involves different mechanisms. Generally, if the repeat function as a protein-binding site, changes in the repeat length can differentially recruit transcription factors at the promoter or splicing factors at the exon-intron junctions. For example, an expansion of repeats can lead to the recruitment of a higher level of binding proteins (i.e. eNOs intron 13). Differential protein recruitment can also be achieved through the formation of alternative length-dependent DNA/RNA structures that ultimately mask or emphasize the protein-binding site (i.e. at the promoter of SLC11A1).

Alternatively, expansions or contractions of the repeat length can indirectly affect gene expression by altering peculiar spacing between other important regulatory *cis*-elements for transcription activation or for exon recognition.

The next pages report few significant examples of polymorphic TG repeats-dependent modulation of gene expression in the context of different cell types.

TG repeats located at the promoter region or 5'UTR

The **hERG** gene, also called KCNH2, is a voltage activated potassium channel involved in a heart condition named drug-induced long-QT syndrome (LQTS). The promoter region of this gene carries a polymorphic TG tract ranging from 20 to 45 repeats among Caucasian and Chinese populations. Real time quantitative RT-PCR analysis on ventricular samples from human subjects showed that longer TG repeats correlates with lower hERG mRNA levels. A luciferase reporter gene carrying the hERG promoter region with different TG

repeats length further confirmed this association. Consistent with lower mRNA levels, endogenous protein levels were also reduced. Lower levels of hERG proteins lead to higher susceptibility to drug-induced LQTS, therefore the TG repeats polymorphism might influence the susceptibility to drug-induced LQTS [129].

The **SLC11A1** gene, also known as NRAMP1 (natural resistance associated macrophage protein 1), has multiple effects on macrophage activation and plays a role in immune response against infections. At the promoter region, SCL11A1 carries a polymorphic TG tract ranging from 4 to 11 repeats. Among the nine alleles carrying a variable number of TG repeats identified to date, alleles 2 and 3 are predominant and exert opposite effects on SLC11A1 expression. Allele 3 has GT9 repeats and drives high SLC11A1 expression (allele 3: T(GT)5AC(GT)5AC(GT)9), while allele 2 contains GT10 repeats and drives lower expression of SLC11A1 (allele 2: T(GT)5AC(GT)5AC(GT)10). Tissue biopsies of allele 3 carriers in fact showed higher SCL11A1 mRNA and protein levels compared to allele 2 carriers [156-158, 170]. These data indicate that the polymorphic TG tract modulates transcription activation in a length dependent manner, through the formation of a Z-DNA structure [74].

The promoter region of the **VWF** gene (Von Willebrand factor) also presents a polymorphic TG element (17 to 23). Although variations in the TG tract do not affect basal VWF promoter activity, they strongly correlate with increased promoter activity upon laminar shear stress. Minigenes carrying longer TG elements lead to higher promoter activity and consequent increase of endogenous mRNA levels in cell cultures. Therefore TG repeats at the promoter modulates the shear stress responsiveness in a length-reliant manner [169].

The **BRN3C** gene, involved in adult-onset hearing loss, carries two distant polymorphic TG tracts in its 5'UTR. These tracts together with a mononucleotide G repeat were shown to co-regulate BRN3C transcription regulation *in vitro* using luciferase reporter constructs. Specifically, TG variability was observed to alter the binding site of the *trans*-acting factor SP1 (stimulating protein 1). The synergic effect of longer TG tracts (TG23) and reduced SP1 levels results in lower transcriptional activity. Alternatively, shorter TG tracts (TG16) had no effect [99]. Similarly, variations in trinucleotide TCC repeat length at the promoter of the EGF gene were found to alter gene expression by affecting SP1 binding [26].

The **STAT6** gene carries a polymorphic TG repeat variant at its 5'UTR ranging from 12 to 17 TG repeats. Luciferase constructs representing different the TG lengths detected in patients differentially regulate the promoter activity *in vitro*. In this case, modulation of the transcriptional activity is not proportional to the repeat length, since only the specific allele TG13 strongly upregulates gene expression [162]. Interestingly, the TG13 allele also specifically associates with asthma, increased production of total serum IgE level and consequent predisposition to allergic disease in British population [163].

TG repeats located in the intronic sequence:

The **beta-ENaC** gene carries a polymorphic TG tract (11, 13 or 14 repeats) in intron 8. Linked with development of essential hypertension, this epithelial channel indirectly modulates blood pressure by regulating sodium and water re-absorption. Patients with hypertension carry longer TG repeats (TG14). A minigene carrying TG14 showed an increased beta-ENaC mRNA expression respect to constructs carrying shorter repeats (11 or 13) [97]. Therefore longer TG repeats are capable of inducing an increase of beta-ENaC mRNA levels

expression and can be linked with higher susceptibility to hypertension. In this case, since the TG repeats are located in intron 8 this effect might be exerted by stabilizing the pre-mRNA rather than by affecting promoter activity.

TG repeats located at the 3'UTR:

The **FGF9** gene is involved in many biological processes including sex development at early embryonic stages in humans. The polymorphic TG repeats at the 3'UTR of the FGF9 gene range from 13 to 16 repeats, with allele TG15 being the most common and allele TG14 associating with gonadal dysgenesis. This association can be due to the fact that TG14 displays stronger promoter activity with respect to TG15 in reporter minigenes. Furthermore, the TG microsatellite also forms structural elements at the 3'UTR affecting mRNA turnover, and the disease-associated TG14 displays longest mRNA stability [118]. Recent mass spectrometry analysis has identified eight *trans*-acting factors binding to the TG15 tract at the 3'UTR of the FGF9 gene, including TDP-43 and FUBP3 (FUSE binding protein 3). Silencing and overexpression of the latter was shown to differentially regulate FGF9 gene expression at the protein level. Therefore, the synergic effect of TG repeats and FUBP3 protein is capable of promoting FGF9 mRNA translation [22].

TG repeats enhance retinoic acid-induced gene repression through the recruitment of PTB:

Retinoic acid regulates cellular differentiation by modulating gene expression and ultimately affects normal embryonic growth and development. In genes susceptible to retinoic acid, the presence of TG repeats in regulatory sequences (introns, 5'/3'UTRs) was shown to strongly enhance the retinoic acid-mediated gene repression. This modulation is granted by TG repeat recruitment of

the *trans*-acting factor PTB (polypyrimidine tract-binding protein). In fact, PTB was reported to bind both DNA single and double-stranded TG repeats sequences and thus modulate retinoic acid-induced repression of gene expression [171].

Consistent with this hypothesis, a microarray analysis of retinoic acid-treated neural crest cells revealed that 41 out of 43 downregulated genes contain TG repeats in either introns or 3'/5'UTRs [171, 172]. Additionally, several genes involved in neuronal crest cells differentiation and embryogenesis that also carry TG repeats, showed significantly lower expression after PTB overexpression. Among them are: Pax7, Otx2, En2 and Snail2 genes. In particular, Snail2 gene carries two TG repeats at its 5'UTR and intron 2-3. Minigene systems carrying various deletions of these two TG tracts showed that the presence of TG repeats is markedly required for retinoic acid-mediated repression of gene expression [171].

UG repeats-mediated splicing regulation

As well as others TRs, UG repeats modulate splicing when located in proximity of splice signals. Presently, four cases of UG repeats-mediated splicing regulation are reported in the literature and reviewed in this section. Besides the splicing outcome, consequences at the protein levels and potential implications in evolution or disease etiology are highlighted. The synergic effect of *trans*-acting factors binding to the UG elements is also presented when uncovered.

UG repeats at the 3'SS of CFTR exon 9: aberrant splicing modulation.

The CFTR gene encodes for a chloride ion channel named cystic fibrosis transmembrane regulator (CFTR). This transporter is located in epithelial cell membranes where contributes to the formation of the mucus. Mutations in the CFTR gene affect the functionality of the channel leading to cystic fibrosis disease.

At the 3'SS of exon 9 a polymorphic locus contains a variable number of dinucleotide TG (ranging from 9 to 13) followed by a T repeat (T5, T7, or T9). Although the length of the polymorphic element varies among ethnic groups, TG12-T7 is considered the most common allele in general population. The increasing number of UG repeats together with the decreasing length of U tracts were shown to correlate with exon 9 skipping in a minigene system [173-175]. *In vivo*, skipping of exon 9 results in a non-functional protein causing monosymptomatic forms of cystic fibrosis [109, 110, 176, 177].

Two different models based on *in silico* and *in vitro* considerations have been proposed to explain how changes in the UG repeats length affect exon 9 splicing outcome and therefore influence disease's incidence. A first model suggests that the UG repeats at the 3'SS form an RNA secondary structure that sequesters splicing signals and inhibits proper exon 9 recognition by the splicing machinery. In a minigene system, UG tracts correlate with increased exon 9 skipping in a length-dependent manner and the substitution of the UG tracts with another dinucleotide repeat (UA repeats) equally contributes to exon 9 skipping [31].

A second model from Baralle's laboratory suggests that the UG element acts a *trans*-acting factor recruiter, and the recruited protein ultimately modulates the splicing outcome. Among the UG binding proteins involved in CFTR exon 9 processing, TDP-43 is the most studied. A detailed description of TDP-43 inhibiting exon 9 recognition in a UG repeats-mediated manner is reported later in the introduction (see page 48) [178].

Besides TDP-43, the potential role in CFTR exon 9 processing played by two CELF family members has also been investigated. Both CELF1 and CELF2 proteins are known to bind CUG repeats and mediate splicing misregulation in muscular dystrophy. Additionally, they display a stronger affinity for UG repeats

elements [179-181]. In a minigene system, CELF1 (also called CUGBP1, CUG binding protein 1) did not seem to affect CFTR exon 9 processing [182]. Concerning CELF2 (also called ETR-3, Elav-type RNA-binding protein 3 or CUGBP2), two independent studies reported discordant data. In a first study, CELF2 enhanced exon 9 inclusion in a minigene system [183]. Alternatively, in a second study CELF2 inhibited CFTR exon 9 inclusion both in a minigene system and in endogenous CFTR mRNA [182]. In this same study authors suggest that structural differences in the minigenes system used can be responsible for the discordant results.

UG repeats at the 3'SS of APOAII exon 3: constitutive splicing regulation

The APOAII gene encodes for the apolipoprotein A-II, a constitutive component of the high density lipoproteins (HDLs) that plays an essential role in lipid metabolism. Additionally, the APOAII protein was recently found to be essential for embryonic development in zebrafish and a crucial factor for nuclear division in mammalian cells [184].

The weak 3'SS of APOAII exon 3 present polymorphic UG repeats, ranging from 12 to 21 with UG19 being the most frequent repeat. Although this UG repeats polymorphism does not affect the APOAII plasma concentrations or the cholesterol levels [89], it represents a peculiar *cis*-acting element for exon 3 constitutive processing. In fact, both deletion and replacement of the UG motif result in exon 3 skipping [185].

The involvement in APOAII exon 3 processing of the *trans*-acting factor TDP-43 binding the polymorphic UG is illustrated later in the introduction (see page 49). No other UG repeats-binding factors have been investigated concerning the regulation of APOAII exon 3 recognition.

UG repeats at the 5'SS of CD36 exon 3: an example of the beneficial effect of instable polymorphic repeats in humans

The CD36 protein (cluster of differentiation 36) is receptor expressed on the surface of various cells types (erythroblasts, macrophages, adipocytes, etc.) that associates with numerous functions (i.e. immune response, hemostasis, lipid transport, etc.). Interestingly CD36 is also capable of influencing the outcome of malaria infection. Malaria is caused by the protozoan parasite *Plasmodium falciparum*, which infects human erythrocytes. The CD36 receptor plays a pivotal role in the adherence of parasite-infected erythrocytes to the capillary endothelia. This adhesion inhibits the immune response to parasite therefore contributing to the pathology of malaria [186].

Genotyping analysis of Thai patients with malaria revealed a polymorphic UG element at the 5'SS of CD36 exon 3, ranging from 11 to 16 repeats [101]. Mild malaria without cerebral involvement specifically associated with the UG12 repeat length, whereas exacerbate cerebral malaria cases associate with longer UG tracts (UG>12). It has been proposed that the allele UG12 confers protection against disease progression by influencing CD36 alternative splicing. Examination of CD36 variants in human blood cells revealed that short UG repeats (UG=11/12) associate with full-length CD36 mRNAs. Alternatively, long UG repeats associate with alternatively splice CD36 isoforms lacking exons 4 and 5. From a functional point of view, this alternative splicing event alters the binding affinity of the parasite-infected erythrocytes for CD36. In fact, CD36 exons 5, together with exons 3 and 6, encode the ligand-binding domain of the receptor protein [187, 188]. On the other hand, the expression of full-length CD36 on macrophages plays a crucial role in the CD36-dependent phagocytosis of the infected erythrocytes [189]. Therefore, the UG12 repeat at the 5'SS of CD36 exon 3 promotes the production of a full-length CD36 protein leading to an efficient phagocytosis and

therefore conferring protection from cerebral malaria [101]. This case represents an example of the beneficial effect of instable polymorphic repeats in human phenotype modulation.

Other disparate polymorphisms in CD36 gene were also found to associate with either protection or susceptibility from severe malaria in African population, in accordance with the key role played by CD36 gene product in this disease [190, 191].

Independent studies in Korean, American, African American and British-Caucasian populations also linked the polymorphic UG repeats in CD36 gene with body mass index and cardiovascular risk (TG=12 leads to higher risk than TG>12) [103-106]. Overall, these studies provide further insight of the favorable role of polymorphic repeats in phenotype modulation. The involvement of *trans*-acting factors has not been investigated in this case.

UG repeats at the 5'SS of SCL8A1 exon 2: tissue-specific alternative splicing modulation

The SCL8A1 (solute carrier family 8, member 1) gene encodes for a plasma-membrane reversible transporter of Na⁺ and Ca²⁺ ions called NCX1 (Na⁺/Ca²⁺ exchanger 1). In humans SCL8A1 contains 12 exons where an unusually long exon 2 encodes for about two-thirds of the entire protein.

Tissue-specific alternative splicing events generate several variants of the NCX1 protein in order to fulfill tissue-specific requirements of Ca²⁺ homeostasis [192]. For example, a short isoform resulting from the circularization of exon 2 is the most abundant in human heart, whereas full-length transcripts have been found in other tissues [193]. The splicing mechanism leading to SCL8A1 exon 2 circularization is unknown, although it has also been found in other genes (i.e. Sry gene and cytochrome P450 gene [194, 195]).

A polymorphic UG tract ranging from 10 to 16 repeats was identified at the 5'SS of SCL8A1 exon 2. No relationship between the length of the polymorphism and cardiac pathologies was detected [161]. However, the UG tract acts as a strong intronic splicing enhancer element and can be involved in the regulation of SCL8A1 expression.

In a minigene system, a short mRNA product is detected in the construct carrying the UG repeats. Alternatively, full-length mRNA is detected when the UG element is deleted or substituted with a non-repeated sequence. Based on these observations, the UG repeat element is capable of modulating alternative skipping of the entire region downstream exon 2, resulting in the production of a short transcript (heart specific) [161].

The mechanisms through which UG repeats mediate alternative splicing as well as potential tissue-specific *trans*-acting factors involved remain to be investigated.

TDP-43

The protein

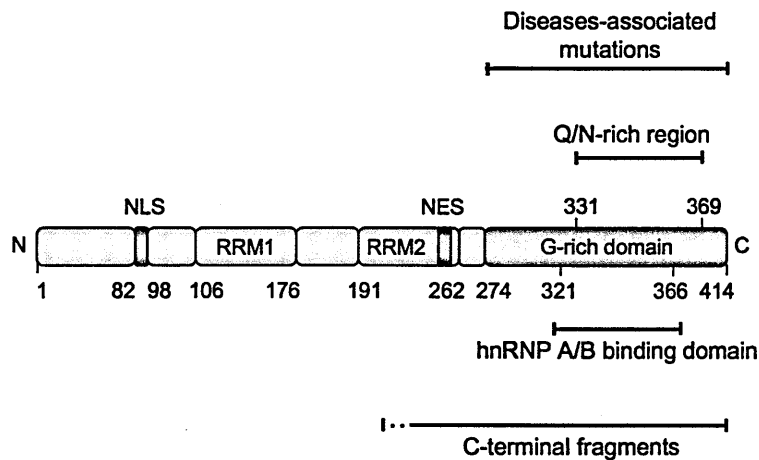
Among the different TG/UG repeats-binding proteins (CELF1, CELF2, TDP-43, PTB, FUBP3, etc.) this dissertation focuses on TDP-43, a protein presenting strong affinity for the TG/UG elements and also involved in numerous neuropathologies that present TDP-43 aggregates.

Highly conserved throughout evolution [196] and ubiquitously expressed [173], TDP-43 is encoded by the TARDBP gene located in chromosome 1. As a member of the heterogeneous nuclear ribonucleoproteins (hnRNPs) family, TDP-43 contains two highly conserved RNA recognition motifs (RRM) flanked by an N-terminal and a C-terminal domain [178] (Figure 2A).

The N-terminal region contains a nuclear localization signal (NLS), consistent with the fact that induced mutations in this region mostly confine TDP-43 into the cytoplasm [197].

Among the two RRM, the RRM1 domain is indispensable for the protein to bind single stranded RNAs and DNAs [173, 198]. The RRM2 domain contains the predicted nuclear export signal (NES) and plays a role in chromatin organization [199] and TDP-43 dimerization [200, 201].

The C-terminal domain mediates protein-protein interactions through a glycine (G) rich tail, which was found to recruit other members of the hnRNP family such as hnRNP A1, A2/B1, and A3 [202, 203]. Moreover, a glutamine/asparagine (Q/N) rich region capable of binding poly-Q aggregates was also identified within TDP-43 C-terminal domain [204]. The strong tendency to aggregate of these Q/N-rich regions and the discovery of aggregated C-terminal fragments in the TDP-43 pathological inclusions suggest a key role for the C-terminal domain in aggregates formation.



- ☐ NLS, nuclear localization signal
- ☐ NES, nuclear export signal
- ☐ RRM1 and 2, RNA-recognition motifs 1 and 2 mutated
- ☐ Glycine-rich domain: hnRNP A/B binding domain (321-366); Q/N rich region (331-369)

Figure 2B: Schematic illustration of TDP-43 domains. The nuclear localization and export signals are reported together with the two RNA binding motifs. At the C-terminal domain both Q/N-rich region and hnRNP A/B binding site are located within the G-rich domain, which is also the region where the majority of the disease causing mutations were detected. Finally, the C-terminal fragments found in TDP-43 aggregates are also illustrated. Fragments have been found to either start at position 208, 219 or 247, whereas the protease involved in the cleavage remains presently unknown.

Localization

In physiological conditions, TDP-43 is ubiquitously distributed in neuronal and non-neuronal cells showing a preferential nuclear distribution, with a weak cytoplasmatic signal [205-208]. Within the nucleus, TDP-43 normally exhibits a fine punctuate pattern on a diffuse nucleoplasmic staining. *In situ* transcription assays demonstrated that TDP-43 is excluded from centromeric and telomeric transcriptionally silent heterochromatin but strongly associated with transcriptionally active euchromatin domains, where specifically colocalizes with perichromatin fibrils (PFs). Enriched in nascent pre-mRNA transcripts and splicing factors, PFs domains are considered active sites of pre-mRNA transcription and cotranscriptional splicing [209] and microRNAs biogenesis [210]. Nonetheless, TDP-43 is absent or rarely detected in transcription-free compartments such as Cajal bodies, perichromatin granules (transient storage sites for mRNAs) interchromatin granule clusters (storage site for splicing factors) and the nucleolus [211]. TDP-43 nuclear localization correlates with its physiological function in different aspects of RNA metabolism.

Additionally, although predominantly nuclear, TDP-43 can shuttle between cytoplasm and nucleus [206]. In fact, a weak cytoplasmatic signal is detected in normal conditions indicating a plausible TDP-43 function also in this cell compartment. This is supported by the observation that, beside its pathological aggregation in cytoplasm of neurons and glial cells [212], TDP-43 has been found to reversibly increase its cytoplasmatic localization in response to neuronal injury in motoneuronal cell lines [213, 214]. In addition, within the cytoplasm of neurons, TDP-43 reversibly associates with RNA transport granules, processing bodies (P-bodies) and stress granules [215-217]. In conclusion, TDP-43 cytoplasmatic localization has been linked with a physiological role both in neuronal response to injury and in RNA transport.

Mis-localization and TDP-43 proteinopathies

Accumulation of intracellular misprocessed and misfolded proteins in the central nervous system is a feature of many neurological disorders. A newly introduced nomenclature defines a group of them as “TDP-43 proteinopathies” based on the presence of TDP-43 as the major pathological substrate for inclusions formation.

TDP-43 proteinopathies include a variety of neurodegenerative diseases with different origin such as amyotrophic lateral sclerosis (ALS) [212, 218], fronto-temporal lobar degeneration (FTLD), poly-Q diseases like Huntington’s disease (HD) and spinocerebellar ataxias (SCA), myopathies and various forms of dementia including Parkinson’s and Alzheimer’s diseases [219]. Among them, the role of TDP-43 in ALS is the most investigated process to this date.

ALS is an adult-onset neurodegenerative disorder characterized by progressive loss of motor neurons that leads to fatal paralysis. Although the cause of the disease remains elusive, the formation of aggregates in motor neurons is a pathological hallmark. Most forms of ALS are sporadic but approximately 10% of the patients have an inherited familial form.

Mutations in several genes have been identified as a risk factor for *familial ALS* such as ANG, SOD1, FUS/TLS, TDP-43 and other rare mutations. Nearly 4% of the familial ALS present TDP-43 mutations, which are mostly localized in its C-terminal domain with the exception of one in the RRM1 domain. Besides one truncating mutation, all the detected variations are missense mutations affecting highly conserved amino acids. Interestingly, cytoplasmatic aggregates of TDP-43 were only detected in familial ALS with TDP-43 and ANG mutations, but not in the SOD1 and FUS/TLS mutated familial forms (reviewed in [219]).

In *sporadic ALS*, mutations in ANG, SOD1, FUS/TLS and TDP-43 were found in less than 1% of the cases. Nevertheless, TDP-43 aggregates have been

found in the totality of the reported sporadic cases. This indicates that TDP-43 mutations (and mutations of the other genes as well) can play an important role in disease pathogenesis but are not the primary causes triggering the formation of TDP-43 inclusions. Consistent with this observation, rare TDP-43 mutations were also identified in patients with FLTD (FTLD-TDP-43) [220-224] but not in other TDP-43 proteinopathies so far.

Within the aggregates, TDP-43 is hyper-phosphorylated, ubiquitinated and abnormally cleaved to generate C-terminal fragments of 20-25 kDa [212, 218]. The pathogenic role of these post-translational modification and cleavage has yet to be established. Phosphorylation of TDP-43 seems to have no effect on aggregation, toxicity or C-terminal cleavage in cellular models [225] but was instead found to correlate with insolubility [226]

The 20-25 kDa C-terminal fragments are predominant in cortical region (brain) whereas full-length TDP-43 inclusions are prevalent in spinal cord [227-229]. However, although the expression in cells of these fragments can recapitulate some pathological features, their pathogenic roles remain unknown.

Ubiquitination of TDP-43 indicates that cellular degradation machineries (such as autophagy and/or the ubiquitin-proteasome system) are involved in TDP-43 aggregates removal. Additionally, TDP-43 affects the expression of HDAC6, a protein associated with aggregation and autophagic degradation of poly-ubiquitinated proteins [230, 231]. Therefore the autophagic system is likely to play a crucial role in the pathogenesis of TDP-43 proteinopathies and of other neurodegenerative disorders characterized by accumulation of protein aggregates (reviewed in [219, 232]).

Although in normal conditions TDP-43 localizes mainly in the nucleus, in patients with cytoplasmatic TDP-43 inclusion a partial clearance from the nucleus is observed [212, 218]. This altered subcellular localization of the protein and the

formation of intracellular aggregates may be critical for the disease pathogenesis. The formation of TDP-43 inclusions may either have a neuro-protective effect through the sequestration of misfolded and truncated species, or a neuro-toxic effect (gain of toxic function). The latter is supported by the fact that TDP-43 overexpression in cellular and animal models leads to increased cytoplasmatic TDP-43 localization and TDP-43 aggregates formation, with consequent cell death or neurodegeneration. However, it has to be noticed that forced synthesis of high levels of TDP-43 does not resemble a physiologically relevant context. In fact, although earlier studies detected an increased amount of TDP-43 mRNA levels in patients with TDP-43 proteinopathies [233, 234], a more recent study found that mRNA levels of TDP-43 were generally normal [223].

Besides the gain of toxic function, sequestration of the protein in cytoplasmatic inclusions with consequent clearance from the nuclei may also result in a loss of normal functions, since TDP-43 may not be available for interactions with protein partners and RNA targets. Additionally, given its role in neuronal recovery, TDP-43 cytoplasmatic localization in neurons has also been suggested to represent a natural response to stress rather than an initiating event in pathogenesis [214].

More hypotheses of how TDP-43 can contribute to neurodegenerative disease pathogenesis in terms of loss or gain of function are described later in the text in association with TDP-43 identified RNA targets and RNA-binding partners. Nevertheless, the exact involvement of TDP-43 misaccumulation in triggering the disease remains to be uncovered.

Binding specificity

Early studies showed that TDP-43 predominantly binds single stranded RNA UG repeats and single stranded DNA TG repeats, and that this affinity increases with the repeat length [178]. In fact, competition assays with synthetic

oligonucleotides showed that TDP-43 binds single-stranded DNA and RNA with TG/UG repeats, but does not bind double-stranded DNA and RNA [173, 178]. Nevertheless, in nitrocellulose filter-binding assays TDP-43 was shown to bind both single and double-stranded DNA, albeit with higher affinity for single-stranded sequences [200]. Also, in hNFL mRNAs, TDP-43 was reported to bind double-stranded RNA where single UG dinucleotides were brought in close proximity to each other by the RNA secondary structure [216, 235].

Consistent with TDP-43 binding specificity for UG repeats is the presence of UG motifs in most of the recently identified TDP-43 RNA targets. Also, an adenine was found to frequently occur in the binding motif of many TDP-43-RNA targets (i.e. (TG)_nTA(TG)_m; more examples in Table 2) [236].

Nonetheless, the growing number of exceptions in TDP-43 binding sequences implies that TG/UG repeats may be the most favored but definitely not exclusive TDP-43 binding site (examples of exceptions include binding sites: in TDP-43 3'UTR [237], in SMN2 exon 7 [238], in SC35 3'UTR [239], in TAR DNA element of HIV-1 in [240], in Nrnx3 and others [86], etc.) (Table 2).

Table 2. TDP-43 binding sequences. Recent CLIP analysis identified TDP-43 clusters in RNA sequences containing UG motifs. These observations point at the UG-rich elements as TDP-43 preferential binding sequence, as predicted in early studies of competition assay with oligonucleotides [178]. Septhon et al. also identified a frequently occurring adenine (A) within TDP-43 binding sites [236]. Numerous examples of non-UG TDP-43 targets were reported by Polymenidou et al. [86] although the exact binding sequence was not indicated. In this table are reported three examples of non-UG binding sites in TDP-43 3'UTR [237], in SMN exon 7 [238] and in TAR DNA element of HIV-1 [240].

TDP-43 Binding Sequences with UG elements	Ref.
(UG)6	[178]
UGUGU	[241]
GUGUG	[241]
UGUA(UG)4	[236]
(UG)2UA(UG)3	[236]
(UG)4UAUG	[236]
(UG)3UA(UG)2	[236]
GUGUA(UG)3U	[236]
G(UG)2UA(UG)2U	[236]
G(UG)3UAUGU	[236]
Non-UG TDP-43 Binding Sequences	Ref.
(GA)3(GC)2GUGC(AG)2ACUU(GGU)2GCAUAA	[237]
(A)5GAAGGAAGG	[238]
C(T)5GCCTGTACTGGG(TC)3TG (DNA sequence)	[240]

RNA targets overview

In order to establish the physiological role of TDP-43 several studies have focused on identifying its targets. Some of them were detected as changes in the splicing outcome in response to altered cell conditions such as TDP-43 silencing or overexpression, including CFTR, APOAII, SC35 and SMN2 [88, 173, 238, 239]. A much higher number of targets (about 400) were predicted from microarray analysis upon TDP-43 silencing. Among them CDK6, HDAC6, and Tbc1d1 have been further investigated and linked with TDP-43 potential physiological functions [199, 242, 243].

Most recently, three independent studies have used CLIP-seq analysis in order to comprehensively identify TDP-43 targets in endogenous conditions [86, 236, 241]. The first study by Sephton et al., has identified 4.352 endogenous TDP-43-RNA-targets in rat brain. The reported targets have been classified in exonic (22%), intronic (45%) or dual (both exonic and intronic) (33%) targets, according to the location of the TDP-43 binding sequence within the genes. TDP-43-RNA-targets are enriched in diverse functional classes such as cell-communication and signaling pathway, RNA metabolism, cell-structure and motility, development, synapses formation and regulation [236]. Strikingly, this enrichment is consistent with the dinucleotide TG repeats distribution within functional classes previously reported in an independent study [20].

The second study by Polymenidou et al., has identified 6.304 endogenous genes targets in mouse brain. In this case, TDP-43 binding sites are mostly intronic (93%); among them 82% locate within 500 bases from the nearest exon-intron boundary. Additionally, splice-junction microarray analysis detected nearly 1000 altered splicing events upon *in vivo* TDP-43 depletion in mouse brain using antisense oligonucleotides. Altered splicing, including both exon inclusion and

exon skipping, was further confirmed for a total of 16 exons using semi-quantitative RT-PCR [86].

In the third and last study by Tollervey et al., mRNA targets of TDP-43 were compared in healthy human brains versus FTLN human brains leading to the identification of 2139 and 2702 targets respectively. The most significant increase in TDP-43 binding in FTLN brains was observed in nuclear paraspeckle assembly transcript 1 (NEAT1) and metastasis-associated lung adenocarcinoma transcript 1 (MALAT1). As reported in the other two studies, TDP-43 binding sites were mostly localized within intronic sequences (60%). In the same study, a splice-junction microarray analysis evaluated splicing changes upon TDP-43 silencing in human neuroblastoma cell lines. A total of 229 altered splicing events were detected and 6 of them have been validated using real time PCR [241].

Noteworthy, a high level of discrepancy was observed among all the studies based on microarray analysis [199, 242, 243]. Similarly, the three reported CLIP-seq analyses not always reported the same RNA targets. This observation might suggest that some of the identified target may be context-specific, cell type-specific or also be subjected to high background noise. However, it has to be considered that the three studies used cells derived from different animals (rat, mouse, human) and that different methodologies were used to deplete TDP-43 (TDP-43 depletion with antisense oligonucleotides mouse brain; human neuroblastoma cell lines silenced with siTDP-43). Possibly these differences account for the discrepancy in the results. Indeed a great amount of work will be required in the future to explore the functionality of all the identified potential RNA targets and to combine all results in a comprehensive meta-analysis.

RNA-binding partners

TDP-43 was found to exist predominantly in high molecular mass complexes associated with RNA. This TDP-43 interactome, where TDP-43 is likely to be present with more than one molecule, has been broadly investigated in several independent studies leading to the identification of numerous TDP-43 interacting proteins [236, 244-246]. Among them are the majority of the known hnRNP proteins (A0, A1, A2/B1, C, D, F, H1, H2, H3, I, K, L, M, Q, R, and U), splicing factors (i.e. SF3a, SF2/ASF), several RNA-binding proteins (i.e. FUS/TLS, CUGBP1, FOX2, ADAR) and components of Drosha microRNA processing complexes (i.e. ILF2/NF43, ILF3/NF90, DDX5, DDX17, DDX3X) [247].

In general, TDP-43 interaction with splicing factors and hnRNP proteins is consistent with its localization in PFBs, nuclear sites of cotranscriptional splicing, and further supports a role for TDP-43 in different aspects of RNA metabolism. Moreover, the interaction with hnRNP family members points at TDP-43 as an integral part of complexes that associate with nascent transcripts and influence their fate. The majority of the others TDP-43 interactors remain to be fully investigated. Nevertheless, in some cases the interaction of TDP-43 with the proteins linked with neurodegenerative disorders has been studied in order to uncover a potentially common pathological mechanism.

Interaction with FUS/TLS

One of the most studied is the interaction between TDP-43 and FUS/TLS (fused in sarcoma/translocation in liposarcoma). These two proteins were found to co-immunoprecipitate in mammalian cells and both TDP-43 G-rich and RRM2 domains are thought to mediate the interaction [244, 246]. TDP-43 and FUS/TLS share several structural and functional features. In fact, like TDP-43, FUS/TLS is an RNA binding protein structurally similar to the hnRNP proteins family with

multifunctional roles in transcription and RNA metabolism. For example, FUS/TLS is implicated in transcription regulation (represses Cyclin-D gene expression [215]), and interacts with RNAPII and the TFIID complex [248]. Also FUS/TLS associates with transcriptional factors, nuclear receptors and Drosha [247] implying a potential role in microRNA processing, mRNA transport and local RNA translation. TDP-43 and FUS/TLS also share an analogous cellular distribution since FUS/TLS is almost ubiquitously expressed and mainly nuclear, with cytoplasmatic accumulation detected in different cell type.

Most interestingly, both proteins are linked to ALS. In fact, in addition to TDP-43 and FUS/TLS mutations, distinct pathological aggregates of TDP-43 or FUS/TLS were found in ALS patients.

Besides ALS, patients with either FTLD or poly-Q diseases also carry FUS/TLS mutation. At present 30 dominant mutations in FUS/TLS gene have been found in nearly 4% of the familiar ALS cases, with the exception of one mutation showing recessive inheritance. Most of them are missense mutations clustered in the G-rich domain and in the extreme C-terminal domain of the protein where the NLS is likely sited [219, 232, 249, 250]. Despite FUS/TLS mutation, patient's neurons and glial cells (brain and spinal cord) present a normal FUS/TLS nuclear staining. However, an abnormal cytoplasmatic and intranuclear accumulation of the protein was also observed. Unlike TDP-43 aggregates, no evidences of biochemical alterations or truncated forms of the protein were detected in FUS/TLS inclusions. Interestingly, FUS/TLS aggregates were found only in neuropathologies where specifically FUS/TLS is mutated (ALS-FUS, FTLD-FUS). Conversely, TDP-43 aggregates were almost ubiquitously present in TDP-43 proteinopathies, mostly in the absence of TDP-43 mutations. The occurrence of TDP-43 or FUS/TLS positive inclusions is generally mutually exclusive, since TDP-43 cytoplasmatic inclusions were not detected in presence of FUS/TLS aggregates. The poly-Q diseases

represent the only exception, where both TDP-43 cytoplasmatic inclusions and FUS/TLS intranuclear inclusions were found in Huntington's disease and SCA type 3 (SCA3) [251]. These observations suggest that FUS/TLS mutations drive a neurodegenerative process that does not depend on TDP-43 mislocalization [229, 249, 250].

Although several studies pointed at similar structure, function and misaccumulation of these two proteins, very little is known on their functional relationship. Only one common RNA target has been identified so far. Specifically, TDP-43 and FUS/TLS possibly co-regulate HDAC6 expression since silencing of both proteins reduced HDAC6 mRNA levels and the simultaneous knock down resulted in a further reduction of gene expression [244].

Interaction with ATXN2

Another interaction with pathological implications is between TDP-43 and Ataxin 2 (ATXN2) [252].

In healthy individuals, the ATXN genes generally carry 22/23 trinucleotide CAG repeats encoding for a long poly-Q tract. As previously reported, expansion of the trinucleotide CAG repeats causes neurodegenerative disorders known as poly-Q diseases (page 9). In the ATXN2 gene in particular, expansion of 34 or more trinucleotide repeats results in SCA type 2 (SCA2) with 100% penetrance. Alternatively, expansions of 27 to 33 repeats are associated with ALS with 5% penetrance. Proteins carrying expanded poly-Q like ATXN2 have a strong tendency to form large cytoplasmatic aggregates with a potentially toxic effect. These aggregates recruit and sequester proteins, which contain poly-Q stretches or alternatively Q/N-rich regions [204].

TDP-43 and ATXN2 co-immunoprecipitate in human cell lines and also co-localize in poly-Q cytoplasmatic aggregates. Additionally, upregulation of yeast

and fly orthologues of the human ATXN2 (Pbp1 and Atx2 respectively) were found to exacerbate TDP-43 toxicity in terms of lifespan shortening and progressive loss of motility [252].

Therefore, TDP-43 interaction with ATXN2 into poly-Q-aggregates might be the cause of TDP-43 translocation from the nucleus to the cytoplasm, with consequent loss of normal TDP-43 function. In fact, TDP-43 sequestration within poly-Q-aggregates was shown to influence the splicing outcome of a reporter minigene carrying CFTR exon 9 [204]. Furthermore, the Q/N-rich region identified at the C-terminal domain of TDP-43 confers to the protein a strong tendency to aggregate. This property identifies a potential self-aggregation mechanism for TDP-43 inclusion formation, as observed for TIA-1 [253]. Co-localization of TDP-43 with poly-Q inclusions in cortical neurons of patients with HD [254] further supports a role for TDP-43 in the pathogenesis of poly-Q diseases.

Interaction with MeCP2, nPTB and CELF1:

TDP-43 was also found to interact with two proteins highly enriched in neurons such as MeCP2 (methyl-CpG binding protein 2) and PTBP2 (polypyrimidine tract binding protein 2) [236].

MeCP2 binds methylated DNAs inducing long-range repression of gene expression. Moreover, it was found to regulate splicing of CD44 exons 4 and 5 in a reporter minigene by interacting with YBX1 (Y box-binding protein 1) [255], a protein that was also found to interact with TDP-43 [246]. Interestingly, MeCP2 function correlates with synapse formation and postnatal development of the nervous system. In fact, loss-of-function mutations or abnormal expression of MeCP2 cause a spectrum of disorders including Rett syndrome, learning disability and autism [256]. Therefore, the interaction between MeCP2 and TDP-43 in a

non-pathological context may imply a functional relationship between these two proteins, which has not been investigated yet.

PTBP2 or nPTB (neuronal PTB) is a member of the hnRNPs family that binds to intronic CU repeats, enhancing the assembly of other splicing factors and RNA-binding proteins [257]. Besides their co-immunoprecipitation ability, nPTB and TDP-43 binding sites are also generally found in close proximity suggesting a possible co-regulation of RNA targets [236]. Similarly, a target co-regulation function is plausible for the interaction between TDP-43 and CEFL1 (or CUGBP1), which share the same UG repeats binding site [182, 246]. Interestingly, CELF proteins and PTB were found to co-regulate splicing in different splicing systems (i.e. alpha-actinin, beta-thropomyosin [258-260]).

Overview of TDP-43 nuclear functions

TDP-43 is a multifunctional protein whose precise role has not been fully elucidated. Numerous studies have provided insight of its implication in several aspects of gene expression regulation including transcription repression, mRNA stability, splicing regulation, RNA transport, translation and microRNA biogenesis. Multifunctional proteins like TDP-43 could have a role in coupling transcription with splicing and other RNA processes. A detailed review of TDP-43 nuclear functions is reported in the following pages.

TDP-43 in transcription

The association of TDP-43 with RNAPII [261] and its co-localization with nascent transcripts in PFs (nuclear site of transcription and cotranscriptional splicing [211, 262]) tend to support a role for this protein in transcription regulation. At present, two cases of TDP-43 exerting an inhibitory effect on transcription are reported in literature.

HIV-1 LTR transcription inhibition

In the first case, TDP-43 inhibits transcription of the HIV-1 viral genome. The HIV-1 gene expression is regulated by several host factors that bind the *cis*-acting sequences located within its long terminal repeat region (LTR). In particular, Tat protein binds a stem-bulge-loop structure called transactivation responsive (TAR RNA) element located at 5' end the HIV-1 mRNAs promoting efficient transcription of viral genes (Tat-induced gene expression) [263, 264]. On the other hand, TDP-43 was found to bind to a pyrimidine-rich motif in the TAR DNA element of the HIV-1 virus (hence the name **TAR DNA binding Protein**), but not to

TAR RNA. Moreover, TDP-43 was proved able to repress both basal and Tat-induced gene expression from the HIV-1 LTR suggesting that its binding to the TAR DNA element might suppress the recruitment of specific transcription factors, ultimately leading to transcription inhibition [240].

The general ability of TDP-43 to bind single stranded pyrimidine-rich sequence was further investigated in Baralle's laboratory highlighting a greatly higher affinity of TDP-43 for single stranded DNA TG repeats. Moreover, as previously shown by Ou et al., TDP-43 did not bind to polypyrimidine single stranded RNAs or to equivalent TAR RNA regions [178].

Acrv1 transcription inhibition

TDP-43 inhibits transcription of the *acrv1* gene, which encodes for the acrosomal sperm protein-10 (SP-10), regulating its spatiotemporal expression during spermatogenesis. In fact, *acrv1* gene remains silent in the somatic tissues and is expressed only in selected stages of spermatogenesis. SP-10 transcription repression is regulated by a molecular mechanism typical of "insulators" [265]. In this mechanism, chromatin-insulators regulate gene expression by reorganizing the DNA in higher-ordered chromatin domains. Protein-insulators bind these domains and physically attach the chromatin fiber to a nuclear peripheral substrate (possibly the nuclear lamina). As a result, in this highly rearranged structure the interaction between enhancer and promoter is blocked and so is transcription. In this case, TDP-43 acts as a protein-insulator by binding on the SP-10 chromatin-insulator [266]. Within the SP-10 insulator, in fact, TDP-43 binds two TGTGTG motifs on the antisense strand, and the disruption of these motifs results in loss of TDP-43 binding capability [267, 268].

Recent ChIP analysis together with molecular markers for pausing of transcription (such as NELF and Ser5s phosphorylation of the RNAPII CTD) revealed that

RNAPII is paused at the *acr1* promoter in correspondence of the TDP-43 binding site. This finding suggests that TDP-43 possibly represses transcription by pausing RNAPII [267]. The attempt to identify the specific TDP-43 domain responsible for transcriptional repression gave discordant results so far. Older studies address the C-terminal domain as the key region for repression [266, 269], whereas a recent report from the same group indicated that the lone RRM1 domain is sufficient for transcription blockage [267].

TDP-43 in general gene expression

Consistent with a gene expression regulatory function, depletion of TDP-43 in cell cultures leads to down- or up-regulation of numerous gene transcripts (about 6000). Among them, TDP-43-mediated regulation of gene expression has been further investigated in three cases: CDK6, HDAC6 and *Tbc1d1* [199, 242, 243]. However, the exact mechanism through which TDP-43 may control HDAC6 and *Tbc1d1* downregulation, and CDK6 upregulation, has not been clarified yet. As both protein and mRNA transcript levels were affected in each case, TDP-43 might either modulate their transcription rate or their mRNA stability.

CDK6

Protein and mRNA levels of CDK6 (cyclin-dependent kinase 6) were found to significantly increase upon TDP-43 removal [199]. These findings suggest that TDP-43 naturally represses CDK6 expression, probably through the binding of the numerous TG repeats within CDK6 gene sequence. Normally CDK6 regulates cell cycle progression by contributing to the phosphorylation and inactivation of the retinoblastoma protein (pRb). When TDP-43 is silenced, CDK6 misregulation results in hyper-phosphorylation of pRb and pRb2/p130, pRb-pathway disruption and subsequent activation of programmed cell death. In fact, genomic instability,

activation of apoptosis and inhibition of neurites outgrowth is observed upon TDP-43 silencing [199, 270]. Additionally, members of the Rho family, known to affect cell morphology and neuronal survival, presented a reduced activity upon TDP-43 silencing [270]. TDP-43 may therefore have a critical role in cell cycle regulation and loss of TDP-43 function may trigger cell apoptosis and neurodegeneration.

HDAC6

Transcript and protein levels of HDAC6 (histone deacetylase 6) decreased upon TDP-43 silencing. Accordingly, a concomitant accumulation of acetyl-tubulin, the HDAC6 major substrate, was also observed [242]. The exact TDP-43-binding site has not been investigated. HDAC6 is implicated in autophagic degradation of poly-ubiquitinated protein aggregates, a cellular defense mechanism that sequesters toxic and misfolded proteins. A potential role in neurodegeneration suppression has also been suggested for HDAC6 [231, 271]. As downregulation of HDAC6 upon TDP-43 silencing results in impaired turnover of aggregating proteins, the consequent accumulation of toxic proteins possibly contributes to the disease onset in TDP-43 proteinopathies [242]. Interestingly, silencing of FUS/TLS (TDP-43 interactor) leads to reduced HDAC6 expression as well. Therefore, both proteins might be required for HDAC6 expression-regulation in normal conditions [244].

Tbc1d1

Whereas *in vitro* and in cell culture studies uncovered a connection between TDP-43 and CDK6 and HDAC6, an *in vivo* study linked Tbc1d1 with Tardbp, the mouse ortholog of the human TDP-43 [243].

Tbc1d1 (tre-2/USP6, BUB2, cdc16 domain family member 1) is an obesity predisposition gene, which mediates glucose uptake upon skeletal muscle

contraction by co-regulating the vesicular traffic of the glucose transporter GLUT4 (glucose transporter type 4). Skeletal muscle levels of Tbc1d1 protein are strongly downregulated in conditional Tardbp-knock down mice. Furthermore, an increased fat metabolism and a lean phenotype were also observed in this animal model. These observations raise the possibility of TDP-43 involvement in energy metabolism and body fat control, consistent with the hyper-metabolic state of some ALS patients. Also, TDP-43-dependent altered fat metabolism may participate in motor neuron degeneration in ALS [243].

TDP-43 in splicing regulation

TDP-43 involvement in splicing regulation is the best-characterized aspect of the protein so far. In different splicing systems TDP-43 either enhances or represses exon recognition through the binding of regulatory elements in proximity of splice signals. Presently, four cases of TDP-43-mediated regulation of RNA processing are reported in literature. In CFTR exon 9 and APOAII exon 3, TDP-43 binds a polymorphic TG repeats element located at the 3'SS [88, 178]. Alternatively in other two cases TDP-43 binds sequences where the TG motifs are absent: in the exonic enhancer in survival of motor neurons 2 (SMN2) exon 7 [238], and in the 3'UTR of the serine/arginine-rich splicing factor 2 (SRSF2 or SC35) [239].

CFTR exon 9

Alternative splicing regulation of the CFTR exon 9 has been extensively studied in Baralle's laboratory over the past ten years [173, 174, 178, 202, 203, 272-276] firstly revealing TDP-43 inhibitory role in exon 9 recognition through the binding of a UG polymorphic sequence at the 3'SS (Figure 3A). Interestingly, monosymptomatic forms of cystic fibrosis were found to correlate with the non-

functional CFTR protein resulting from exon 9 skipping in the final CFTR transcript [176]. Consistent with TDP-43 inhibitory effect, higher rates of exon 9 skipping were observed upon TDP-43 overexpression, whereas increased exon 9 inclusion was detected upon TDP-43 silencing [178, 277]. These results were validated both in exogenous minigenes and endogenous mRNA from patient's lymphoblasts [272]. Most recently, sequestration of TDP-43 in poly-Q aggregates also results in increased CFTR exon 9 inclusion in a minigene system [204].

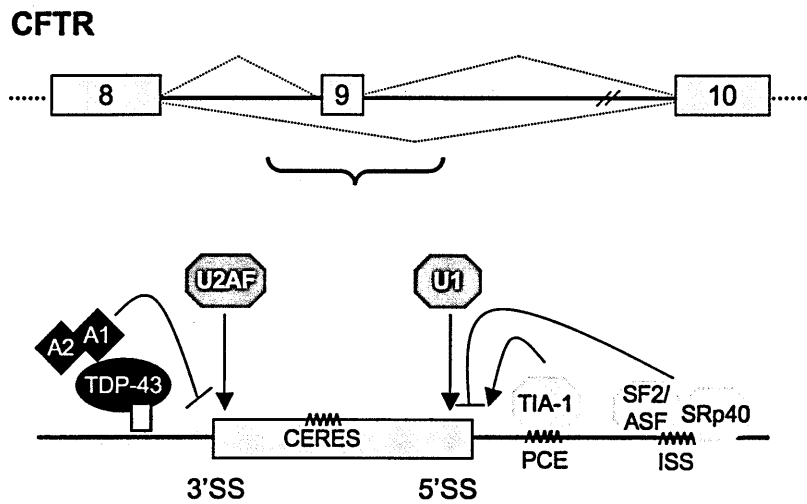
The interaction of the C-terminal domain of TDP-43 with other hnRNP family members was shown to play a synergic role in exon 9 processing both in mouse and in minigene systems [196, 202, 203]. In addition to the UG repeats-mediated recruitment of TDP-43, further *cis* and *trans*-acting elements located at the 5'SS and within the exon were found to influence exon 9 inclusion levels in the final CFTR mRNA (Figure 3A). In fact, at the 5'SS, the TIA-1 splicing factor (T-cell-restricted intracellular antigen-1) promotes exon inclusion through the binding of a pyrimidine-rich enhancer region (PCE); this interaction is thought to stabilize the U1 snRNP interaction with exon 9 weak 5'SS [278]. In addition, an intronic splicing silencer (ISS) within intron 8-9 recruits SF2/ASF and SRp40 downregulating exon 9 recognition. Although SR proteins generally promote exon recognition, in this splicing system SF2/ASF and SRp40 were found to favour exon 9 skipping [175, 276]. Finally, composite regulatory elements (so called CERES - composite exonic regulatory elements of splicing) located within the exon 9 sequence present context-dependent enhancing or silencing properties, which further affect exon 9 processing [273]. Although numerous modulators of exon 9 processing have been identified, TDP-43 plays a dominant inhibitory role that overrides other regulatory elements. In fact, TDP-43 silencing resulted in exon 9 inclusion even in presence of mutations that disrupt PCE, and CERES elements and in SF2/ASF overexpressed conditions [272].

APOAII exon 3

The APOAII exon 3 is constitutively spliced in the final mRNA and no mutations or diseases have been associated with exon 3 skipping so far. In this case, TDP-43 has an evolutionary relevant role in exon 3 splicing regulation. Just like previously observed for CFTR exon 9, TDP-43 binds the polymorphic UG repeats at the 3'SS of exon 3 exerting an inhibitory effect on exon inclusion (Figure 3B) [88]. In addition, in intron 2-3 an ISS composed of mononucleotide G repeats recruits hnRNP H1 and further promotes exon 3 skipping. However, constitutive exon inclusion is guaranteed by the presence of an ISE at the 5'SS of exon 3 that specifically recruits the SR proteins SRp40 and SRp55 [279] and of an exonic splicing enhancer (ESE) within exon 3 that binds SF2/ASF and SC35 (Figure 3B) [279].

Based on the evolutionary conservation of all these elements, it has been proposed that these two enhancer sequences evolved in order to offset TDP-43 and hnRNP H inhibition of exon 3 inclusion. Consistently with this theory, and similarly to CFTR exon 9 system, TDP-43 silencing overrides the need of other *cis*-acting element since constitutive inclusion of exon 3 occurs even with disrupted or mutated intronic and exonic enhancers [88].

A



B

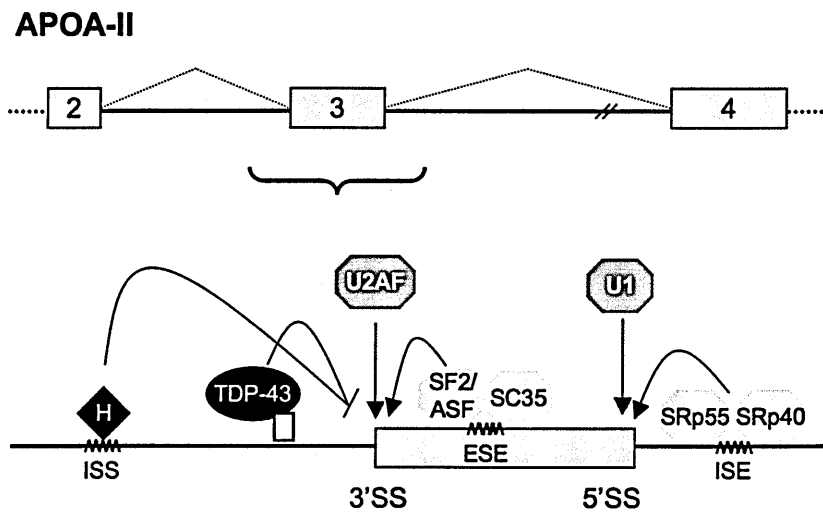


Figure 3. TDP-43 involvement in 3'SS definition. *Cis*-elements and *trans*-acting factors involved in the processing of CFTR exon 9 (A) and APOAII exon 3 (B). The *trans*-acting proteins are illustrated as different shapes whereas the *cis*-elements are represented as zigzag lines. The red colored rectangle represents the UG repeats bound by TDP-43. An enhancing effect on exon inclusion is illustrated as an arrow, an inhibitory effect with a "—|".

There are two nearly identical copies of survival of motor neuron (SMN) genes in the human genome: SMN1 and SMN2. The critical difference between them is a C (in SMN1) → T (in SMN2) change at position 6 of exon 7 (+6C>T). This transition results in predominant skipping of exon 7 in the SMN2 transcripts, leading in an unstable, rapidly degraded protein. Mutations or deletion of the SMN1 gene result in spinal muscular atrophy (SMA) because SMN2 gene fails to compensate impaired SMN production [280-282].

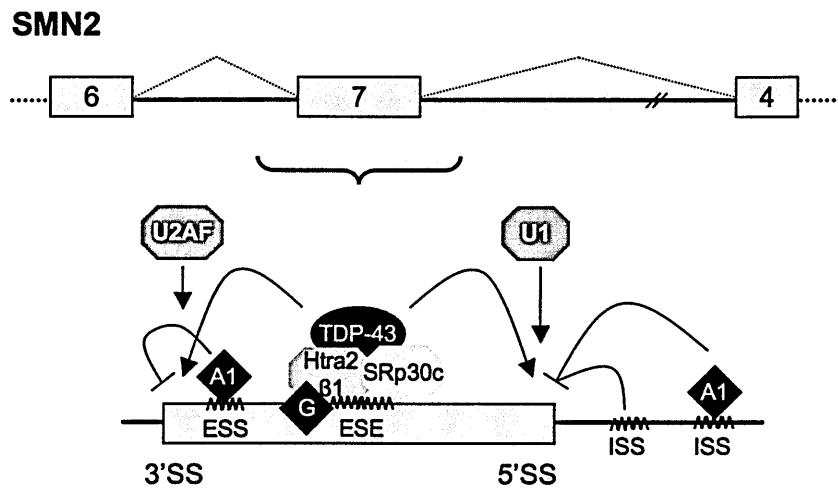
Given its implication with SMA, an autosomal recessive disorder with degeneration of the alpha-motor neurons, the SMN2 exon 7 alternative splicing has been extensively studied over the years leading to the identification of several *cis*- and *trans*-acting elements involved (Figure 4A). The +6C>T change creates a strong exonic splicing silencer (ESS) that recruits hnRNP A1 [283], but also disrupts a pre-existing SF2/ASF binding site [284]. At the 5'SS of the exon 7 there is another hnRNP A1 binding site [285] and an additional ISS was also found in intron 7-8 [286]; both sequences further promote exon 7 skipping. This strong inhibition of exon 7 recognition is balanced by the presence of an ESE in the central region of SMN2 exon 7. This enhancer recruits several factors such as SRp30c, hnRNP G, hTra2-β1 and TDP-43, which act as positive regulators of exon 7 inclusion [238, 287-289]. Specifically, the integrity of an AG-rich sequence (AAAAAGAAGGAAGG) within the ESE was proved essential for the assembly of a multimeric complex. Overexpression of TDP-43 in the SMN2 exon 7 minigene system resulted in an increased exon 7 inclusion, indicating that TDP-43 can enhance exon recognition in a TG repeats-independent manner [238]. On the other hand, TDP-43 silencing failed to affect SMN2 exon 7 processing.

SC35 3'UTR and auto-regulation

Several proteins involved in RNA processing, including all SR proteins family members and some hnRNPs, were shown to negatively auto-regulate their own expression to maintain homeostatic levels (i.e. PTB [290]; hnRNP L [291], SF2/ASF [292], SR proteins [293-295]). This negative feedback loop mechanism generally involves unproductive alternative splicing events such as alternative inclusion of poison-cassettes-exons containing a premature termination codon or alternative retained introns. When this happens, the resulting mRNAs are targeted for degradation through non-sense mediated decay (NMD) or exosome decay.

TDP-43 was shown to be involved in the auto-regulating mechanism of the SC35 splicing factor [239]. As observed for other members of the SR proteins family, the overexpression of SC35 results in a significant decrease of endogenous SC35 mRNA levels, indicating an auto-regulatory capability. Specifically, SC35 promotes alternative splicing with both poison-cassette-exon and intron retention events, in addition to alternative polyadenylation events in its 3'UTR. The resulting mRNAs are unstable and get degraded through NMD [296]. A key role in the SC35 auto-regulating mechanism is played by a 60-nucleotides enhancer region in SC35 exon 5 (terminal untranslated exon at the 3'UTR), which recruits SC35 protein leading to the splicing of the upstream intron (intron 4-5) and consequent formation of unstable mRNAs (Figure 4B) [239]. Beside SC35, a complex pattern of proteins was found to binds to this 60-nucleotides enhancer region, including TDP-43. It has therefore been suggested that TDP-43 competes with SC35 for this binding site, leading to inhibition of intron 4-5 processing. In fact, TDP-43 overexpression resulted in poor exon 5 3'SS recognition and consequent inhibition of intron 4-5 processing in a minigene system.

A



B

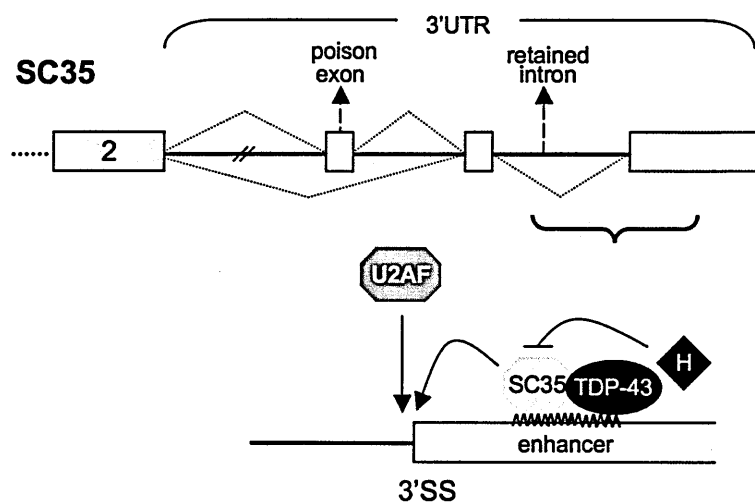


Figure 4. Examples of TDP-43 involvement in splicing regulation. *Cis*-elements and *trans*-acting factors involved in the processing of SMN exon 7 (A) and SC35 exon 5 (B). The *trans*-acting proteins are illustrated as different shapes whereas the *cis*-elements are represented as zigzag lines. An enhancing effect on exon inclusion is illustrated as an arrow, an inhibitory effect with a “—”.

A recent study in Baralle's laboratory has shown that TDP-43 as well can control its protein level by self-regulating its transcript expression [237]. This auto-regulatory activity is exerted through the binding of several low affinity binding sites in its 3'UTR, which do not resemble the canonical UG repeats high-affinity binding site.

Unlike other auto-regulating proteins described in literature, TDP-43 self-regulation does not seem to involve either alternative splicing or alternative polyadenylation events, but regulates the translation efficiency of its RNA transcripts by affecting the mRNA stability. In this case, the exosome macromolecular complex was found to be partially involved in the degradation of the unstable mRNAs, whereas NMD played a marginal role. Downregulation of the endogenous TDP-43 expression observed upon overexpression of an exogenous TDP-43 is also supported by *in vivo* data in brain lysate from a mouse model [297, 298].

Consistent with these data, a study aimed to identify TDP-43 target sequences revealed an enrichment of reads in the 3'UTR of TDP-43 itself [236]. Interestingly, the abundance of TDP-43 target sequences at the 3'UTR of other genes suggests that TDP-43 may similarly control post-transcriptional regulation of numerous other transcripts other than SC35 and its own.

Additional evidences of TDP-43 involvement in splicing regulation

Recently, a splice-junction microarray study from Polymenidou et al. has identified 965 altered splicing events, including 203 alternative splicing variations, upon TDP-43 depletion using antisense oligonucleotides in adult mouse brain [86]. Some of these identified targets have been validated using semi-quantitative RT-PCR and real time PCR highlighting alternative inclusion of 9 exons and alternative skipping of 7 exons. Among them, exon 18 of the receptor for

progranulin Sortilin1 was found alternatively included upon depletion of TDP-43 [86].

Similarly, another study by Tollervey et al. identified 158 alternative splicing events and 71 constitutive splicing changes in human neuroblastoma cell lines depleted of TDP-43 [241]. In this case as well some targets have been validated with semi-quantitative RT-PCR and real time PCR. As a result, TDP-43 knockdown was shown to modulate alternative exon inclusion in four cases: i) in *ZNF92* exon 2 where a UG repeated sequence is located at the 3'SS of exon 2; ii) in *REEP6* exon 5, which carries a UG-rich element in the intronic downstream sequence; iii) in *CACNA1C* exon 13 that carries a short UG elements in the downstream sequence as well; and iv) in *PILRB* exon 25 carrying UG repeats in the upstream sequence.

On the other hand, depletion of TDP-43 resulted in alternative exon skipping in *CDK5RAP2* exon 39 and *GPBP1L1* exon 4, both carrying UG repeats in the downstream region [241].

TDP-43 in microRNA biogenesis

In addition to TDP-43-mediated splicing regulation, a potential role for TDP-43 in microRNAs biogenesis has been proposed. This alternative nuclear function is supported by TDP-43 association with several components of the Drosha microRNA-processing complex such as ILF2/NF43, ILF3/NF90, DDX5, DDX17 and DDX3X [247, 299], and by TDP-43 localization in nuclear regions linked with microRNA processing such as PFs and P-bodies [211]. Most interestingly, specific changes in the total microRNA population have been reported in TDP-43 depleted cell cultures [300]. Among the differentially expressed microRNAs, TDP-43 was shown to directly bind both let-7b microRNA sequence and let-7b hairpin precursor in correspondence of a UG-rich motif GGU(UG)3GUU. Downregulation

of let-7b upon TDP-43 depletion resulted in the upregulation of four different let-7b microRNA targets (VAMP3, STX3, DYRK1A and LAMC1) [301]. Another microRNA, miR-663, was instead upregulated upon TDP-43 silencing, leading to the downregulation of its target EPHX1 (epoxide hydrolase 1, xenobiotic). In this case, TDP-43 was shown to bind the pri-miR-663 precursor sequence GUC(UG)2U, but not the microRNA sequence itself, which does not present UG motifs [300]. In correlation with the growing evidences of microRNAs implication in neurodegenerative diseases [302], the up-mentioned microRNAs targets (VAMP3, STX3, DYRK1A, LAMC1 and EPHX1) can all be somehow linked with different neuropathologies.

Overview of TDP-43 cytoplasmatic role

Cellular functions of TDP-43 extend beyond its nuclear localization because TDP-43 can shuttle to the cytoplasm [206] through the importin-beta nucleocytoplasmic transporter [213]. An important but yet to be fully characterized TDP-43 cytoplasmatic function has been proposed specifically for neuronal cells. Within the cytoplasm of neuronal cells, in fact, TDP-43 was found to reversibly associate with several cytoplasmatic RNA granules, such as RNA transport granules (Staufen and FMRP positive) and RNA-processing bodies (XRN-1 positive), pointing at TDP-43 as an active player in RNA transport and translation. Moreover, TDP-43 co-localization with stress granules (TIA-1 positive) places this protein as a potential participant in cellular response to stress [216, 303]. Finally, misregulation of TDP-43 cytoplasmatic functions may contribute to neurodegeneration in a still undetermined manner.

TDP-43 in RNA transport granules

Synthesis of specific proteins in spatially restricted areas and in response to temporally specific signals is achieved through a precise mRNA localization and mRNA translational control. This mechanism has been observed in both somatic and neuronal cells. In the latter, neuronal activity represents the specific signal capable of triggering the recruitment of translationally dormant mRNAs to the dendritic spines and their local translation. For example, dendritic localization of BDNF and Trk-B mRNAs increase after neuronal activity [304]. Likewise, synaptic stimulation was shown to deliver beta-actin and CaM-KII-alpha mRNAs in dendrites and promote in-site synthesis of the corresponding proteins [305-307]. These mRNAs are transported to the dendrites within large granules also called

“localizing complexes”, which include RNA-binding proteins, ribosomal elements and translational factors [308]. Interestingly, localizing complexes also include microRNAs that inhibit local translation of their associated target mRNAs. In fact, in response to neuronal activity microRNA precursors are degraded, microRNAs are no longer generated and the translational repression is defeated [309, 310].

The assembly of the localizing complexes on the mRNAs occurs co-transcriptionally and then undergoes dynamic remodeling at different stages [311]. The mRNAs targeted for dendrite localization carry a *cis*-acting dendritic localization sequences generally located in their 3'UTRs. Protein factors bind to these 3'UTR dendritic localization sequences and subsequently mediate the microtubule-based transport. Moreover, given that localizing mRNAs are translationally silent during their transport, the protein factors also act as translational repressors by targeting different regulators of the translation process (reviewed in [312, 313]). Among the multifunctional protein factors are ZBP1 and ZBP2 (zipcode-binding protein 1/2), Staufen, CPEB (cytoplasmic polyadenylation element-binding protein) and FMRP (fragile X mental-retardation protein).

TDP-43 co-localization with FMRP and Staufen proteins in neurons together with its RNA-binding activity and the presence of a strong nuclear export signal support a role for TDP-43 in this nuclear to cytosolic shuttling of mRNAs and ultimate regulation of their translation. In accordance with this theory, TDP-43 granules were detected in dendrites of neuronal cells in culture and the number of granules increased after depolarization by KCl stimulation. Therefore, TDP-43 actively translocates to the dendritic spines in response to neuronal activity in order to transport specific mRNAs [215]. Within the granules, TDP-43 was found to associate at least with two mRNAs: beta-actin and CaM-KII-alpha (calmodulin-dependent kinase II-alpha) [215], whose mRNA levels were previously known to increase upon depolarization leading to local synthesis [314]. Furthermore, TDP-

43 was shown to act as a local translation repressor in cultured neuronal dendrites, possibly through its reported interaction with elongating factors and ribosomal proteins [215].

Altogether these studies indicate that TDP-43 targets dendritic mRNAs and regulates their transport and local translation. At present, a well-characterized mRNA cargo carried by TDP-43 is the hNFL mRNA. Specifically found in neurons, hNFL (human neurofilaments light chain) are cytoskeletal proteins members of the intermediate filaments family. TDP-43 binds the hNFL mRNAs at their 3'UTR, likely in correspondence of single UG dinucleotides brought in proximity to each other by RNA secondary structure [216, 235] mediating its stability. Other two proteins involved in ALS pathogenesis, 14-3-3 protein and the mutated SOD1, were shown to co-regulate hNFL mRNAs stability together with TDP-43 [216]. Interestingly, NF aggregates were observed in ALS patients and their formation may be triggered by a misregulation of their transport by TDP-43.

Besides TDP-43, other members of hnRNP family like hnRNP K, E1 and E2 bind hNFM (human neurofilaments medium chain) mRNAs through a CA-rich motif and mediate mRNA stability transport and ultimate translation [315, 316].

TDP-43 in P-bodies

Within the cytoplasm, TDP-43 was also found to co-localize with P-bodies. Therefore, it has been speculated that TDP-43 might transport mRNAs to P-bodies as well in order to inhibit their translation upon neuronal activation [215]. Degradative granules or P-bodies, in fact, are known to mediate mRNA degradation, RNA storage and micro-RNA mediated repression of translation [317]. However, further studies are needed to fully elucidate this functional aspect of TDP-43.

TDP-43 in stress granules

Stress granules (SGs) are cytoplasmatic dense aggregation of proteins and RNAs that appear upon stress signals in order to protect the untranslated mRNAs from harmful conditions. The untranslated mRNAs molecules are stored in stalled translation pre-initiation complexes and can either be degraded or re-initiate translation when SGs gradually dissolve after neuronal recovery (reviewed in [318]).

TDP-43 was found to reversibly associate with SGs in response to oxidative insult [303], heat shock [217] and induced neuronal injury (axotomized motor neurons) [213, 214]. In disagreement with earlier studies [303], TDP-43 was recently found to co-regulate both assembly and maintenance of stress granules in response to oxidative stress [217]. In addition, TDP-43 can differentially regulate two SGs nucleating proteins, either at protein and mRNA levels. Specifically, TDP-43 upregulates TIA-1 and downregulates G3BP (Ras-GAP SH3 domain binding protein), two newly identified targets for TDP-43. However, the specific mechanism through which TDP-43 regulates their transcript and protein level has not been investigated so far. Interestingly, two ALS-linked mutant TDP-43 were found to alter stress granule dynamics. These loss-of-function mutations with regards to SGs formation (R361S and G348C) localize in the C-terminal domain of TDP-43 indicating a possible involvement of this region in SGs interaction [217, 319]. Like TDP-43, other members of the hnRNP family such as hnRNP A1, A2 and Q redistribute to the cytoplasm and localize with SGs under specific stress conditions [217, 320].

Aim of the project

Splicing of both alternative and constitutive exons is co-regulated by the presence of short intronic or exonic sequences (*cis*- elements) acting as splicing enhancers or silencers. Among them, the dinucleotide UG repeats located at the 3'SS of an exon were found to inhibit splicing of the downstream exon through the recruitment of the *trans*-acting factor TDP-43. A present, this inhibitory effect has been validated in two splicing systems: the CFTR exon 9 alternative splicing and the APOAII exon 3 constitutive splicing [88, 173].

In addition, it has been observed that, when located in proximity of the 5'SS of an exon, the dinucleotide UG repeats can promote the alternative skipping of the downstream exons, leading to shorter mRNA products. This effect has been highlighted in two splicing systems: the SCL8A1 exon 2 and the CD36 exon 3 [101, 161]

Based on these previous observations, the aim of this project has been to investigate whether the UG repeats located at the 5'SS of an exon might influence the processing of the upstream exons. Using both artificial and natural systems, the impact on splicing of three different variables has been investigated: i) the length of the UG repeats; ii) the distance of the UG repeats from the 5' end of the exon; iii) the strength of the 5'SS (in terms of consensus value) of the upstream exon; and iv) the presence of UG repeats-binding proteins.

Given its great affinity for the UG repeats elements, the involvement of the splicing factor TDP-43 in this regulatory process has also been studied.

Materials and methods

Constructs preparation

Reporter minigenes used

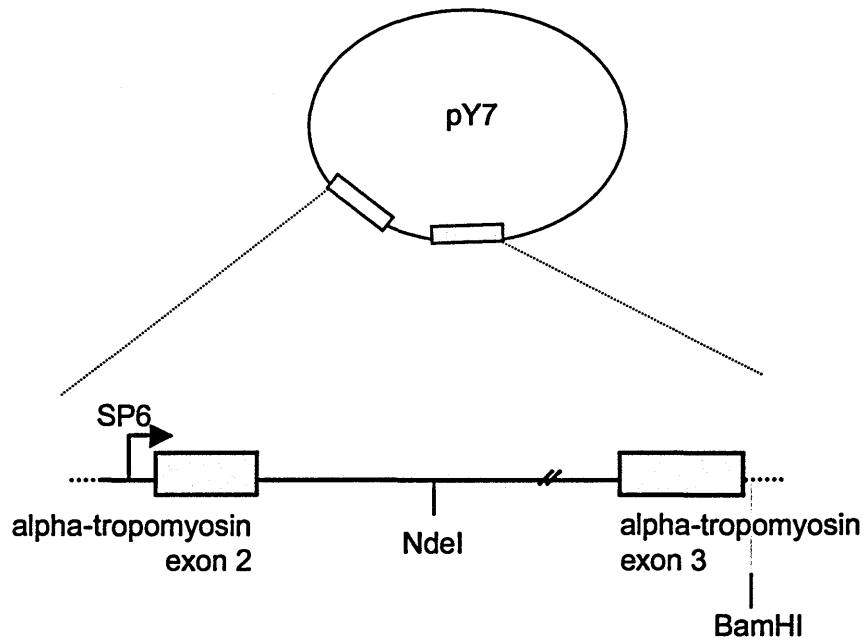
In this study, the splicing pattern of several exons has been investigated by using two reporter minigenes/splicing-constructs (pY7, pTB) inserted in different plasmids backbones (pBMN, pBluescript and pcDNA3).

The splicing-construct pY7 contains the alpha-tropomyosin exons 2 and 3 separated by a 111-nucleotides synthetic intron sequence with a 5'SS partially derived from the human beta-globin gene. The SP6 promoter controls the expression of this reporter minigene that is inserted into the pBMN vector backbone [321] (Figure 5a).

The splicing-construct pTB is composed by the alpha-globin exons 1, 2 and 3, and fibronectin exon 3 under the control of the SV40 promoter within the pBluescript vector backbone [322] (Figure 5b).

In addition, in this work the pY7-based splicing-constructs were inserted into the pcDNA3 plasmid, which carries the CMV promoter (pY7-pcDNA3, i.e. Figure 6d, e, etc.). This insertion has been made because the SP6 promoter in the pY7-pBMN plasmid is not suitable for the in cell culture and the *in vitro* coupled transcription-splicing assays required for this study.

a pY7 reporter minigene (pBMN backbone)



b pTB reporter minigene (pBluescript backbone)

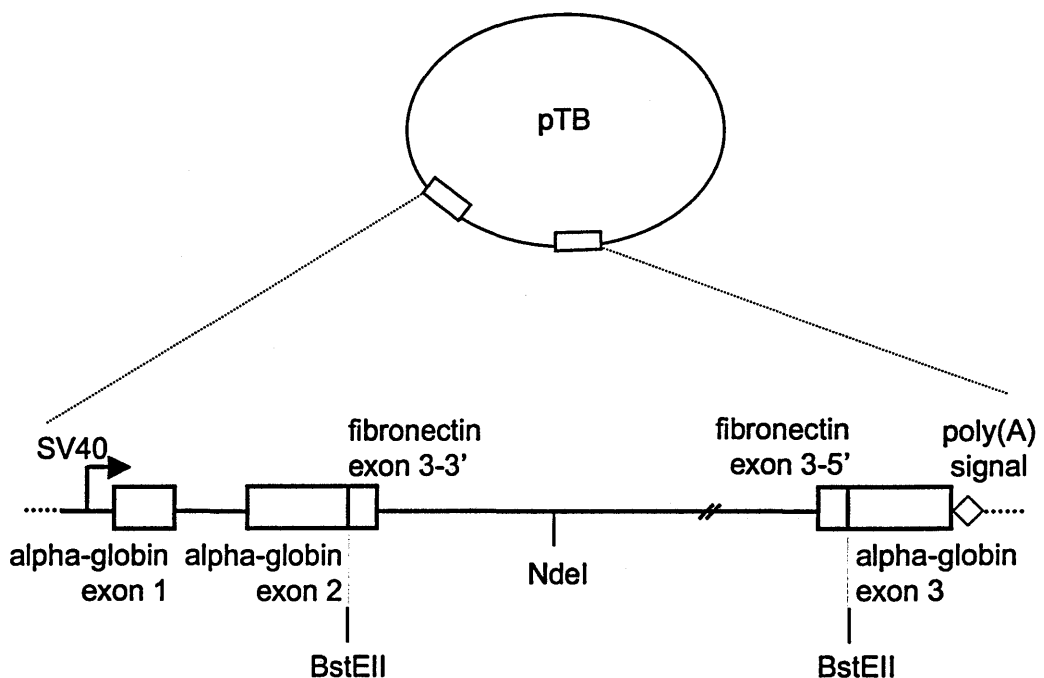


Figure 5: Schematic representation of the reporter minigenes used in this study. a) py7 reporter minigene; b) pTB reporter minigene. The exons defining the minigenes and the cloning restriction sites are indicated.

Constructs: TG6 at 8, TG6 at 18, TG12 at 8, TG12 at 18 in the pY7 reporter minigene

Four constructs with different number of TG repeats (TG6 and TG12) at two different distances (8 and 18 nucleotides) from the 5'SS were constructed by modifying the downstream region of the alpha-tropomyosin exons 2 in the pY7 reporter minigene (Figure 6a, b). The DNA fragments containing the exon 2 modified 5'SS were generated by PCR reaction using the pY7 minigene reporter as template. In example, for the TG6 at 8 nucleotides construct two distinct PCR have been performed, one using the oligos TG6 at 8-sense and pY7-Xbal-antisense and another one using the oligos pY7-KpnI-sense and TG6 at 8-antisense. Corresponding oligos were used for the other constructs (oligos used for PCR reactions: TG6 at 8, TG6 at 18, TG12 at 8, TG12 at 18-sense and antisense and the external oligos pY7-KpnI-sense and pY7-Xbal-antisense; oligo sequences are reported in appendix in Table 4).

Afterwards, a mix of the two amplified fragments was used as template in another PCR reaction using the external oligos on the alpha-tropomyosin exon 2 and 3 carrying the KpnI and XbaI enzymes restriction sites respectively (oligos: pY7-KpnI-sense and pY7-XbaI-antisense). Finally, the obtained DNA inserts and the pY7 reporter minigene were digested with the KpnI and XbaI enzymes (New England Biolabs), gel extracted (EuroClone) and ligated with T4 DNA ligase (New England Biolabs) according to manufacturer's instructions.

Constructs BRCA1 5'SS exon 12 wild type and mutant in the pY7 reporter minigene

Using the same cloning strategy described above, the 5'SS of alpha-tropomyosin exons 2 in the pY7 reporter minigene was substituted with the 5'SS of the BRCA1 exon 12 in the two constructs BRCA1 exon 12 wild type (WT) and

mutant (MUT) (oligos: BRCA1 5'SS WT-sense and antisense; BRCA1 5'SS MUT-sense and antisense; Table 4) (Figure 6c). The mutation in the BRCA1 MUT construct is a pathogenic mutation found in breast cancer patients, known to alter the splicing outcome in the final BRCA1 mRNA *in vivo*.

Insertion of the pY7 reporter minigenes in the pcDNA3 vector

The six pY7-based splicing-constructs (four TG repeats constructs and two BRCA1 WT and MUT constructs) and the unmodified reporter minigene (TG0 construct) as well, were inserted into the pcDNA3 plasmid, which carries the CMV promoter suitable for *in vitro* coupled transcription-splicing and in cell culture experiments. To do so, external oligos on the alpha-tropomyosin exon 2 and 3 carrying the KpnI and XbaI enzymes restriction sites respectively were used in PCR reactions using each of the seven different constructs as templates (oligos: pY7 KpnI-sense and pY7 XbaI-antisense; Table 4). The DNA inserts generated were then cloned into the pcDNA3 plasmid using the KpnI and XbaI restriction sites (Figure 6d, e, f, g).

Additional BRCA1 constructs: BRCA1 5'SS and BRCA1 exon 12

Both BRCA1 5'SS WT and MUT in the pY7 reporter minigene have been also inserted in the pTB reporter minigene (Figure 6h). To do so, external oligos on the alpha-tropomyosin exon 2 and 3 carrying the BstEII enzyme restriction site were used in a PCR reaction using the constructs in figure 6g as templates (WT and MUT) (oligos: pY7-BstEII-sense and pY7-BstEII-antisense; Table 4). Afterwards, the DNA inserts generated were cloned into the pTB vector using the BstEII restriction site.

Using the same cloning strategy, BRCA1 5'SS WT and MUT in the pY7 reporter minigene have been also inserted in the pTB reporter within an additional

exon cassette (hERG exon 7) using the BglII restriction site (oligos: pY7-BglII-sense and pY7-BglII-antisense; Table 4) (Figure 6i).

The BRCA1 constructs carrying the whole length exon 12 were generated by PCR amplification of the DNA fragment from genomic DNA (control-WT and patient-MUT). Subsequently, the fragments were cloned into both the pTB and pY7 reporter minigenes using external oligos carrying the NdeI restriction site (Figure 6j, k). The pY7-based BRCA1 exon 12 constructs were also inserted into the pcDNA3 vector using KpnI and XbaI restriction sites (Figure 6l).

Constructs: EPHX1, SGK2, NTRK1, RXRG, ETF1 and Δ TG ETF1 in the pTB reporter minigene

Each DNA fragment (containing exon of interest and flanking intronic sequences) was amplified from genomic DNA from HeLa cells using oligos carrying the NdeI enzyme restriction site at their extremities (Table 4) (EPHX1 exon 5, SGK2 exon 7 and 8, NTRK1 exon 12, RXRG exon 7, ETF1 exon 7) and cloned into the pTB reporter minigene (Figure 6m, n, o, p, q).

The Δ TG ETF1 construct was generated through modification of the downstream sequence of ETF1 exon 7 construct. Using the ETF1 minigene as a template, a first set of PCR reactions was performed (oligos: Δ TG ETF1-sense + ETF1 exon 7-antisense and Δ TG ETF1-antisense + ETF1 exon 7-sense; Table 4). The DNA fragments were gel extracted and used as templates for a second PCR reaction (oligos: ETF1-sense + ETF1-antisense). The DNA insert obtained was then cloned into pTB vector using the NdeI restriction site (Figure 6r).

The generated constructs were sequenced using the CEQ 2000XL sequencer (Beckman). Afterwards large scale preparations of plasmidic DNAs were performed using JetStar purification kit (Genomed) according to the manufacturer's instructions.

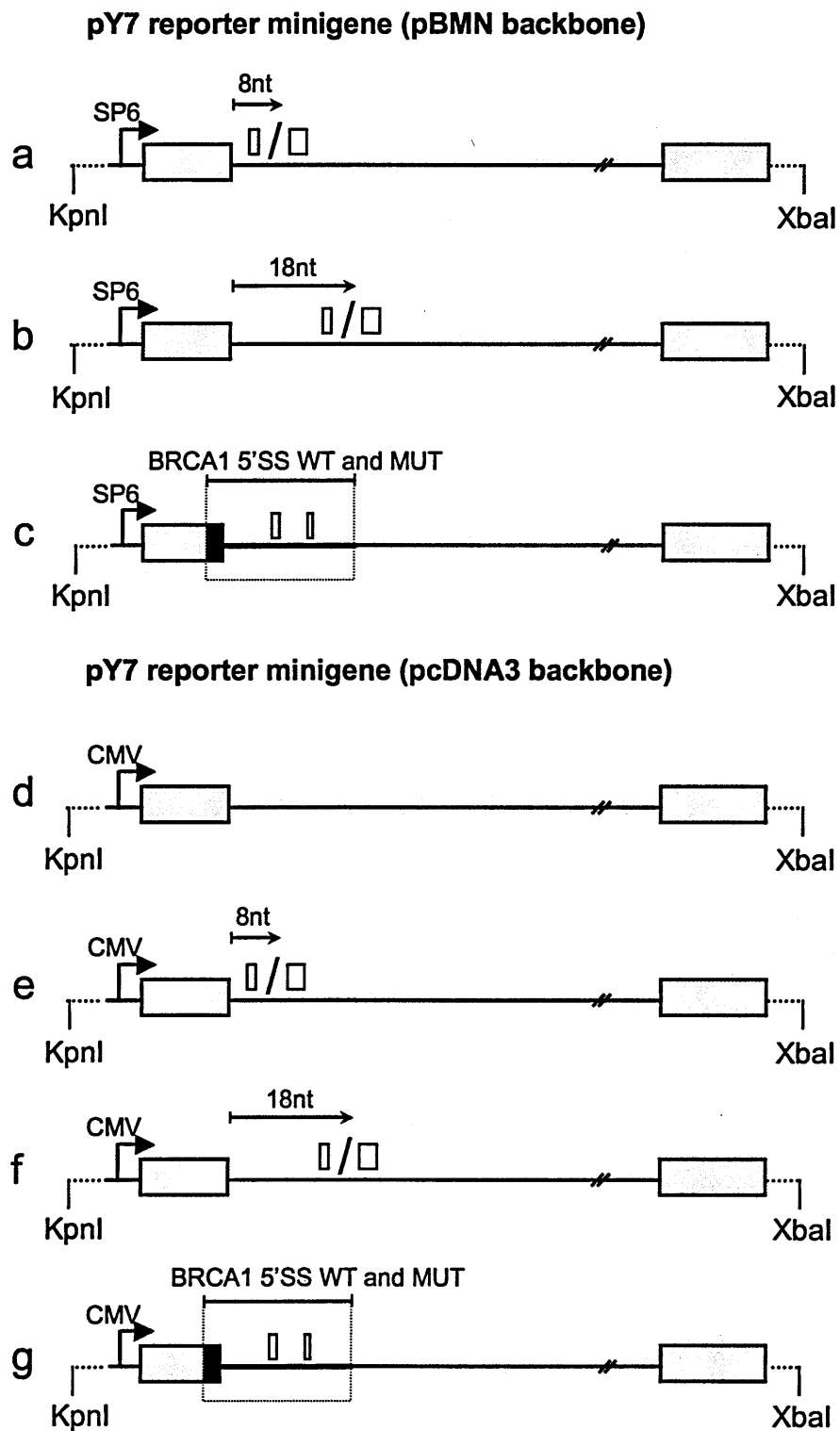


Figure 6a-g. Schematic representation of the constructs generated in this study. The red rectangles represent the TG repeat elements; the size of the rectangles is proportional to the length of the repeat. In figures 6a, b, e and f the two red rectangles separated by a "/" are representative of the insertion of either TG6 or TG12 elements. In figures 6c and g the red rectangles represent the two natural GT6 and GT4 at the 5'SS of the BRCA1 exon12, WT and MUT.

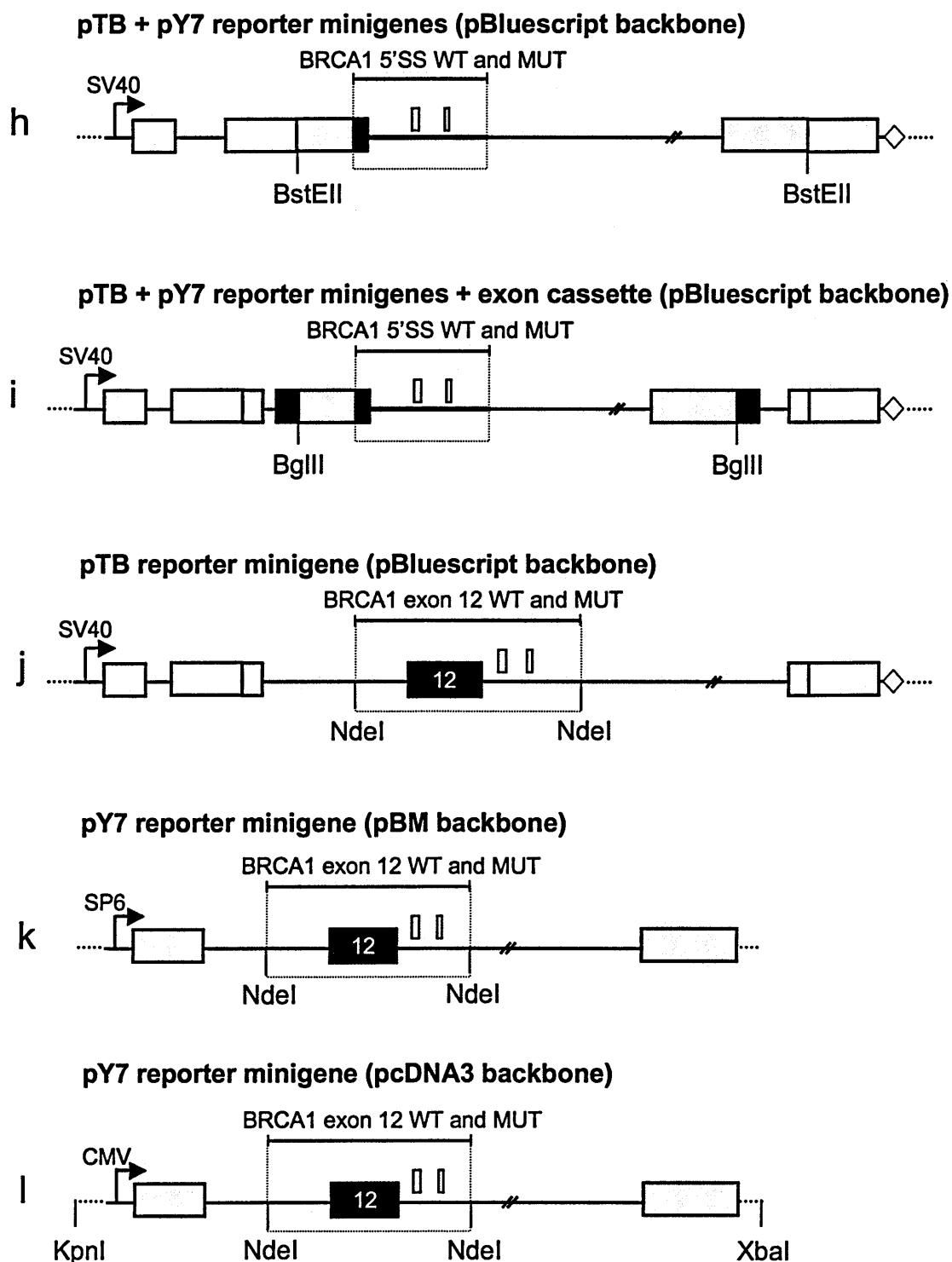


Figure 6h-l. Schematic representation of the constructs generated in this study. The red rectangles represent the TG repeat elements; the size of the rectangles is proportional to the length of the repeat. In figures 6h-l the red rectangles represent the two natural GT6 and GT4 at the 5'SS of the BRCA1 exon12, WT and MUT.

pTB reporter minigene (pBluescript backbone)

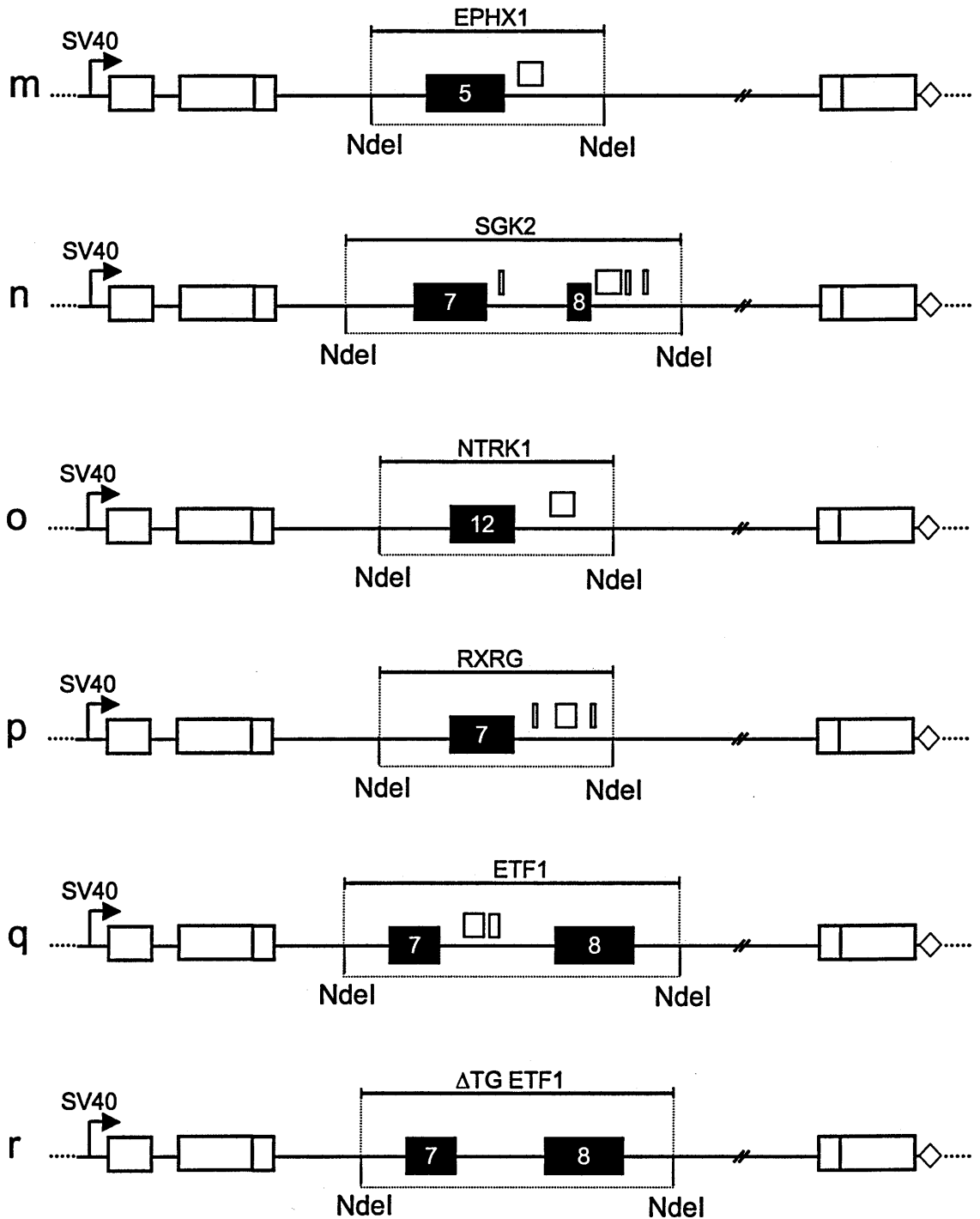


Figure 6m-r: Schematic representation of the constructs generated in this study. The red rectangles represent the TG repeat elements; the size of the rectangles is proportional to the length of the repeat.

TDP-43 constructs

In cell culture experiments, four different recombinant TDP-43 were used: i) TDP-43 wild type (WT); ii) TDP-43 with domain RRM1 and RRM2 mutated, thus deprived of its RNA-binding activity (RRM1,2M); iii) TDP-43 lacking the entire C-terminal domain (Δ C); iv) TDP-43 lacking the hnRNP binding site and the Q/N rich region (Δ 321-366) (Figure 7). All the TDP-43 constructs were previously engineered using the pGEX-3X vector as reported in references: [173, 198, 203, 206].

TDP-43 constructs:

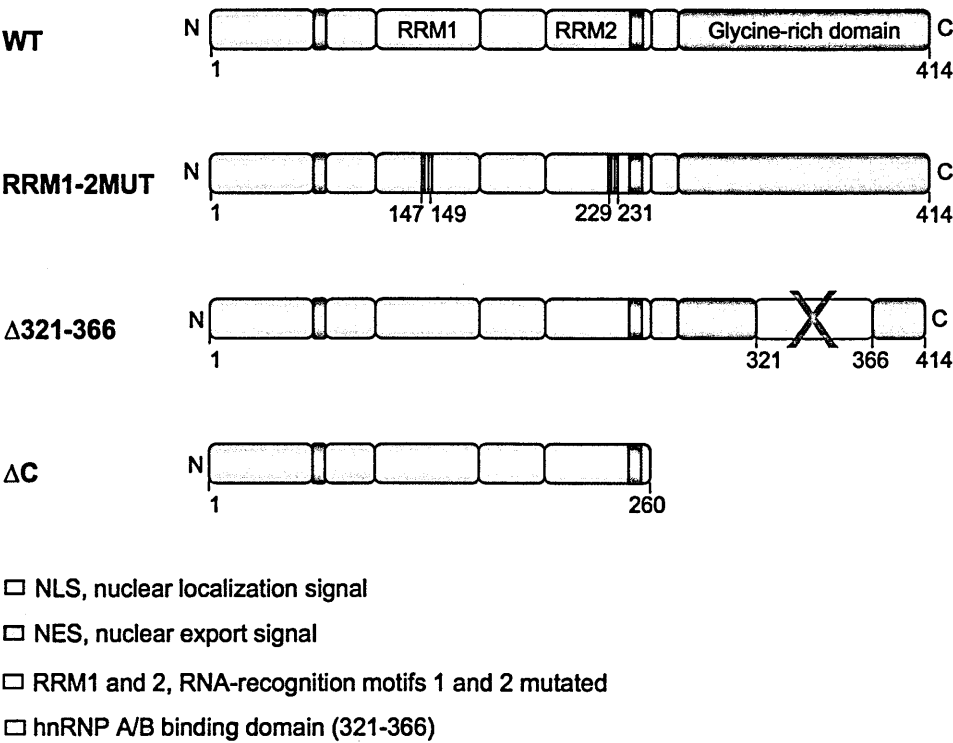


Figure 7. Schematic representation of the TDP-43 constructs used in this study. Four recombinant siRNA-resistant and flag-tagged TDP-43 proteins are illustrated in this picture. The different domains of the protein are represented with different colors. In the RRM1,2 MUT TDP-43 are reported in red color the four mutated phenylalanines at the indicated amino acidic positions. In the Δ 321-366 construct the deletion of the 321-366 part is illustrated in a light green color and crossed with an "X".

***In vitro* traditional splicing assay**

The pY7 reporter minigenes containing the inserted exons of interest were first linearized by digestion BamHI enzyme (Promega) and then phenol/chloroform extracted and EtOH precipitated. The resulting DNAs were then transcribed into pre-mRNAs using the SP6 RNA polymerase (Promega). The 40µl-volume reaction consisted of: 2µg of DNA, 1X transcription buffer (Promega/Stratagene), 7.5mM ATP, CTP and UTP, 1.5mM of GTP, 5mM m7G CAP analog and 1U of SP6 RNA polymerase (Promega). When producing hot RNAs, 1.5mM of UTP instead of 7.5mM was used and 1µl alpha-32P-UTP (800 Ci/mmol) (Amersham) was added. Transcription was allowed at 37°C for 2 hours; then the samples were treated with 1U of RNase-free DNaseI (Roche) for 15 minutes. Thereafter, the SP6-transcribed pre-mRNAs were phenol/chloroform extracted, EtOH precipitated and resuspended in RNase-free water.

Subsequently, splicing reactions were preformed at 30°C for 2 hours in 25 µl reaction mixtures containing 3.2mM MgCl₂, 500µM ATP, 20mM creatine phosphate, 2.7% w/v polyvinyl alcohol, 20mM HEPES pH=7.3, 15µl of standard HeLa nuclear extract (CilBiotech) and 2µg of SP6-transcribed pre-mRNA.

When necessary, 0.4 µg of GU6 or (GU)2A6(GU)2 RNA oligonucleotides (Sigma, Table 4) were pre-incubated at RT for 10 minutes prior addition to the splicing reaction.

The processed RNAs were then extracted from the reaction mix using acid phenol/chloroform and then EtOH precipitated.

Hot RNAs were directly re-suspended in RNA loading buffer, denatured at 96°C for 3 min and separated in denaturing polyacrylamide-urea gels with proper concentrations. The gels were then dried and the RNA fragments visualized by phosphor imager (Molecular Dynamics) or detected by autoradiography.

Cold RNAs were retro-transcribed and analyzed by RT-PCR as described later in the text (see “cDNA synthesis and analysis” at page 75).

***In vitro* coupled RNAP II transcription and splicing**

The DNA fragments were PCR-amplified from the pY7-based minigenes inserted in pcDNA3 (oligos used for PCR reactions: CMV-sense and BGH-antisense; Table 4). Afterwards, the resulting DNA fragments were phenol/chloroform extracted and EtOH precipitated.

In vitro transcription-splicing reactions were performed as described in [261] at 30°C in 25µl reaction mixtures containing 0.5mM ATP, 3.2mM MgCl₂, 20mM creatine phosphate, 1µl alpha-³²P-UTP (800 Ci/mmol), 15µl standard bulk nuclear extract and 0.6µg of DNA. Alpha-amanitin (10ng/25µl reaction) was used to block transcription. At each time point, an aliquot of the reaction was removed from the coupled RNAPII transcription-splicing reaction, de-proteinized, phenol/chloroform extracted, EtOH precipitated and analyzed on a denaturing polyacrylamide-urea gels as mentioned above.

***In vitro* splicing assay in cell culture**

The pTB and pY7-pcDNA3 reporter minigenes carrying the exons of interest were transfected in HeLa cells in a liposome-mediated manner. Specifically, a total of 3X10⁵ HeLa cells were seeded in 35mm Petri dishes and grown in DMEM-GlutaMAX-I media (GIBCO) supplemented with 10% fetal bovine serum (EuroClone). After 24 hours, the 80% confluent plates were transfected with 1µg of plasmidic DNA using the effectene transfection reagent (Qiagen) according to the manufacturer's protocols.

Finally, total RNAs were collected after 48 hours using the TRIzol reagent (Invitrogen) according to manufacturer's instructions, treated with 1U of RNase-

free DNase I (Roche) at 37°C for 15 minutes, phenol/chloroform extracted and EtOH precipitated. The RNA quantity was detected using the Nanodrop spectrophotometer instrument (Thermo Scientific) and equal amounts were used for cDNAs synthesis (see “cDNA synthesis and analysis” later in the text).

When needed, depletion of TDP-43 was performed by RNA interference. In this case, 8×10^4 HeLa cells were seeded in 35mm Petri and grown for 24 hours. For each 35mm Petri dish, 1.2 nmol of siTDP-43 (Dharmacon) or of the controls (non-targeting siRNA and si-Luciferase; Dharmacon) were transfected using the oligofectamine reagent (Invitrogen) according to the manufacturer's protocol in Opti-MEM GlutaMAX-I media (GIBCO). After 24 hours siRNAs transfections were repeated to ensure proper TDP-43 depletion, given the abundance of this protein in the cells. Six hours after the second siRNAs transfection cells were washed with PBS1X (137mM NaCl, 2.7mM KCl, 10mM Na₂HPO₄, 1.8mM KH₂PO₄, pH 7.4) and 1µg of the plasmidic DNAs (minigenes) was transfected using the effectene transfection reagent as described above.

When the overexpression of recombinant TDP-43 was required, 1µg of the plasmidic DNAs carrying the different siRNA-resistant and flag-tagged TDP-43 sequences (Figure 7) was simply added to the transfection mixture used for the minigene construct.

Cells were finally harvested after 48 hours and total RNA was extracted as mentioned above. When siRNA treatment and recombinant TDP-43 overexpression were performed, total proteins were also extracted using the TRIzol reagent according to manufacturer's instructions.

cDNA synthesis and analysis

Both the RNAs extracted from the cells and the RNAs from the *in vitro* traditional splicing assay were retro-transcribed into cDNAs and subsequently

PCR-amplified. The 12µl-volume reaction mixtures composed of 200ng of random primers (Pharmacia), 1mM dNTPs mix (Promega) and 2µg of RNA were denaturated at 94°C for 3 minutes and then chilled on ice. Afterwards specific solutions (Invitrogen) were added to the reaction mix: 10mM DTT, 1X first strand buffer (10mM Tris-HCl pH8.4, 50mM KCL, 2.5mM MgCl₂), 2U of MML-V reverse transcriptase enzyme (moloney murine leukemia virus). The final mixture was then incubated at 37°C for 1 hour and 2µl of the retro-transcribed cDNAs were used as PCR template (oligos used for PCR reactions: pY7 exon 2-sense and pY7 exon 3-antisense for the pY7-based constructs; alpha 2,3-sense and bra-antisense for the pTB-based constructs; Table 4). All RNAs were tested for DNA contamination by retro-transcription PCR (RT-PCR) lacking the reverse transcriptase enzyme.

Finally the PCR products representing the splicing products were separated on 2% agarose gels and ethidium bromide-stained. Quantification of band intensities in the acquired pictures was performed using the ImageJ64 software (available at <http://rsb.info.nih.gov/ij>).

Western blots and antibodies

Upon quantification with Bradford assay, 30µg of proteins from each sample were mixed with 10X protein loading buffer (20% w/v SDS, 1M DTT, 0.63M Tris-HCl pH 7, 0.2% w/v bromophenol blue, 20% v/v glycerol, 10mM EDTA pH 7). Conventional gels SDS-PAGE were performed using the required percentage of polyacrylamide (37,5:1 acrylamide:bis-acrylamide; ProtoGel National Diagnostics), run at 40mA in 1X SDS-PAGE running buffer (50mM Tris, 0.38M glycine, 0.1% w/v SDS). SDS-PAGE gels were then blotted on standard nitrocellulose membrane and incubated in blocking solution (PBS1X, 0.2% Tween, 2% milk) over night. Primary antibodies were incubated with the membrane in the same buffer for 2 hours. Afterwards, the membrane was then washed three times

in washing solution (PBS1X, 2%Tween), incubated with the proper HRP-secondary antibodies (anti-mouse/rabbit, Dako) for 1 hour, washed twice and stained with ECL reagent (Thermo Scientific). Finally, an autoradiography was taken on Kodak Biomax XAR films.

Endogenous TDP-43 was detected with anti-TDP43 polyclonal antibodies [178] (1:2000); recombinant TDP-43 with anti-FLAG M2 monoclonal antibody (1:2000) (Sigma F1804); and the loading control with anti-tubulin monoclonal antibody (1:5000) (Sigma T5168).

Spliceosomal assembly analysis

The SP6-transcribed pre-mRNAs were used in a traditional splicing reaction with the same conditions described above (see “*In vitro* traditional splicing assay” page 73). Prior addition to the reaction mix, nuclear extracts were incubated at RT for 10 minutes with the GU6 or (GU)2A6(GU)2 RNA oligonucleotides (Table 4). Aliquots were collected at 5 and 30 minutes time points and reaction was stopped by addition of 2.5µl of 5X spliceosome gel loading buffer (1X TBE, 30% v/v glycerol, 0.5g/l Heparin, 0.1% w/v bromophenol blue). The splicing complexes were separated through electrophoresis using a 4% polyacrylamide native gel run at 300V for 3 hours. Finally, gels were dried and visualized by phosphor imager (Molecular Dynamics).

***In silico* analysis and predictions**

The bioinformatic analysis of the dinucleotide TG repeats located at the 5'SSs was performed by downloading the “exon/intron boundaries sequences” file from the “alternative splicing database” (ASD) combined with the “alternate transcript database” (ATD), which contains alternatively transcribed/spliced exons in nine different organisms collected from literature (available at

<http://www.ebi.ac.uk/asd/>) [323, 324]. Firstly, the TG repeats with a repeat unit number equal or greater than six were simply highlighted in the Microsoft Word downloaded file. Secondly, the exons carrying the TG repeats sequences at the 5'SS were selected and identified one by one based on the reported accession code using the “nucleotide database” in the NCBI website (available at <http://www.ncbi.nlm.nih.gov/nuccore>).

The 5'SS consensus values (CVs) were calculated according to the “splice site prediction by neural network” website.

(available at http://www.fruitfly.org/seq_tools/splice.html).

The comparative genomic alignments for the ETF1 and RXRG exons 7 and the identification of the different alternative spliced transcripts for all the genes listed in Table 3 were detected using the “Ensembl” website.

(available at http://www.ensembl.org/Homo_sapiens/Info/Index).

In silico analysis of the ESE and ISE sequences were performed using the “human splicing finder” website.

(available at http://www.umd.be/HSF/4DACTION/input_SSF).

Restriction reports indicating all the enzyme restriction sites present in a sequence were predicted using “the sequence manipulator site” website.

(available at <http://www.bioinformatics.org/sms/>).

CLIP analysis

The CLIP analysis (Cross Linking and ImmunoPrecipitation) is a procedure used for the identification of protein-binding sites on RNA, in the context of an intact cell. This methodology is often mentioned in the thesis introduction and discussion as an alternative approach pursued from other laboratories to provide genome-wide insight of the TDP-43 potential binding sites and RNA targets.

Furthermore, in the process of selecting potential TDP-43 targets, we also used CLIP data from Prof. J. Ule's laboratory (personal communication) that reported specific TDP-43 binding sites within the RNAs sequences of ETF1, EPHX1, NTRK1, RXRG and SGK2. Briefly, in a CLIP analysis cells are dissociated and UV-treated on ice leading to the formation of a covalent bond between protein and RNAs. After cell lyses and partial RNA digestion, the protein of interest is immunoprecipitated together with the bound RNA fragments. Prior to SDS-PAGE and transfer to nitrocellulose membrane, RNAs are ligated to the 3' end with an RNA linker and radioactively labeled at the 5' end. Afterwards, the membrane is exposed to X-ray film, and the region of the membrane corresponding to protein-RNA complexes is excised, proteins are digested and a 5' end RNA linker is ligated to RNAs. The CLIP tags are then amplified by RT-PCR, using DNA primers with sequence complimentary to the RNA linkers, and finally sequenced.

Statistical Analysis

A statistical analysis was performed on the investigated groups of data using the *Student's t-test* (also known as two-sample t-test), the most common statistical data analysis procedure for hypothesis testing. Specifically, the tested hypothesis was the "null hypothesis" (the hypothesis of "no difference" between two measured phenomena/groups of data) versus and the "alternative hypothesis" (the hypothesis of significant difference between two measured phenomena). For example, in this work for each minigene the alternative hypothesis tested has been whether there was a significant (not due to chance) difference in exon inclusion between groups A and B if group A was treated with a no-target siRNA and group B with a TDP-43 siRNA, or alternatively if group A was treated with a

wild type recombinant TDP-43 and group B with a mutated form of TDP-43.

The p value (probability value) represents the estimated probability of rejecting the null hypothesis and accepting alternative hypothesis. Therefore, if the p value is lower than the chosen significance level, the null hypothesis is rejected and the analyzed data are considered “statistically significant”. Although the choice of significance level is arbitrary, in this case the conventional 5%, 1% and 0.1% levels were used ($p < 0.05$ *; $p < 0.01$ **; $p < 0.001$ ***).

In order to perform this test, all the experiments were repeated at least three times since the *Student's t-test* investigates the null hypothesis of no difference between the means of two normally distributed groups of data.

Results

Influence of UG repeat length and distance from the 5'SS on exon processing

UG repeats lengths

In order to investigate whether the length of a UG element located near the 5'SS of an exon can affect the exon processing, four constructs with different UG repeat lengths were engineered using the pY7 reporter minigene (Figure 8a). This minigene [321] has been previously used in Baralle's laboratory to prove that UG repeats at the 3'SS inhibit processing of the downstream exon. Furthermore, in this case the inhibition was proved to be TDP-43-mediated and to correlate with the UG repeat length [277]

For this study, the downstream region of the pY7 alpha-tropomyosin exon 2 was modified by adding either TG6 or TG12 repeats at 18 nucleotides from the 5' end of the exon (Figure 8a). This distance from the 5'SS was chosen because close enough to the 5'SS to affect the splicing outcome, but also far enough to not directly interfere with U1snRNP recruitment. The unmodified minigene, reported here as TG0, does not carry TG repeats at the 5'SS and represented the negative control for these experiments. The TG0, TG6 and TG12 at 18 nucleotides constructs were then used for *in vitro* traditional splicing assay. The RNA was transcribed using the SP6 enzyme and then allowed to splice for 60 and 120 minutes using commercial nuclear extract. The splicing intermediates and products (pre-mRNAs and mRNAs) were retro-transcribed, PCR amplified and separated in 2% agarose gels.

As shown in Figure 8B, no significant differences in splicing efficiency were

detected at 60 or 120 minutes between the TG0 control and the two constructs carrying 6 or 12 TG repeats at 18 nucleotides from the 5' end of the exon (Figure 8B). Earlier time points of 30-60 and 45-90 minutes were also tested but reported no differences in splicing patterns (data not shown).

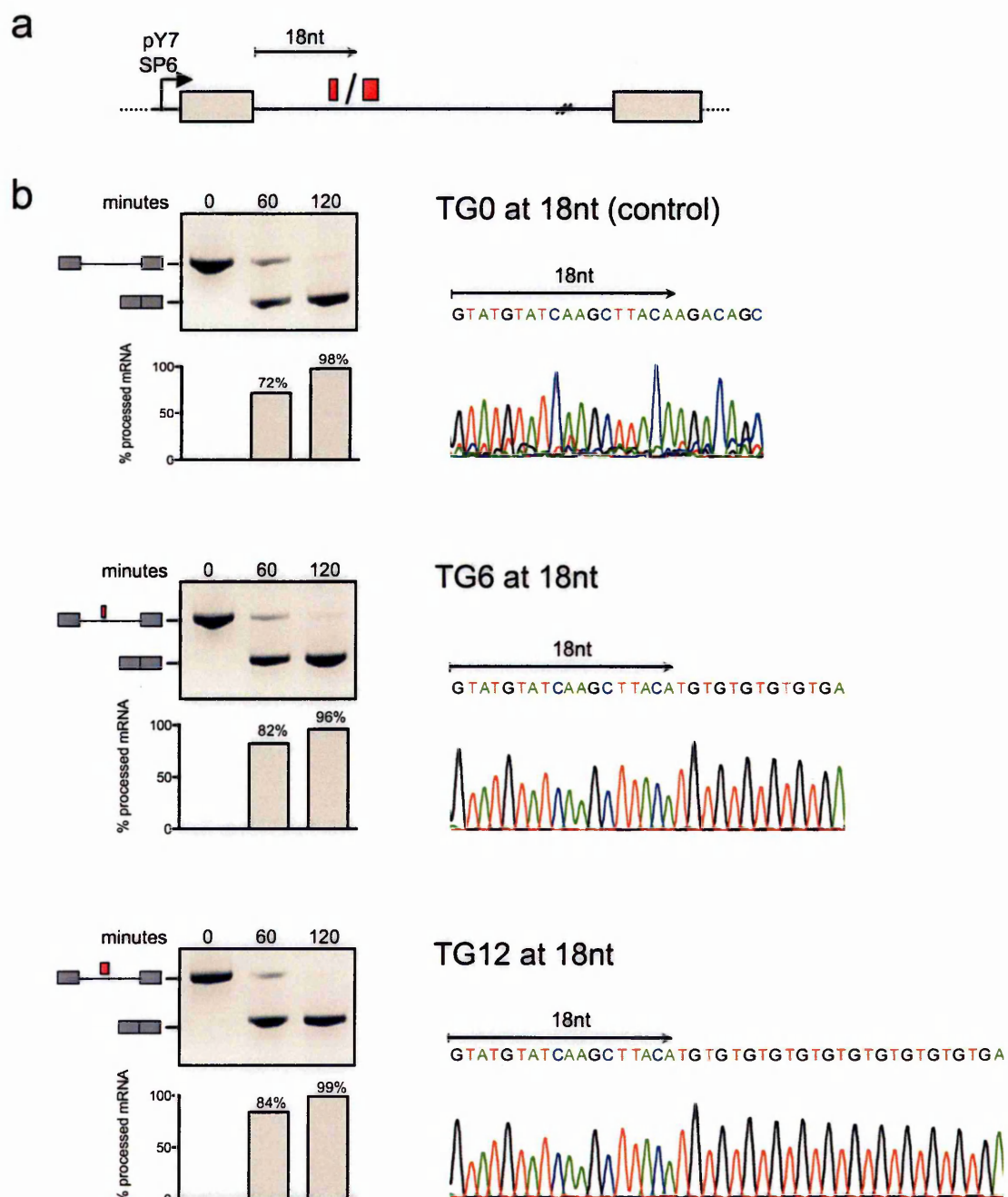


Figure 8. Insertion of TG6 and 12 at 18 nucleotides from the 5' end of exon 2 does not affect exon 2 processing. a) Schematic representation of the minigenes used. b) Splicing assay of pY7 constructs carrying TG repeats at 18 nucleotides from the 5' end of the exon 2 and the TG0 control. For each construct (TG0, TG6 and TG12), is reported the splicing pattern, the quantification of the bands and the sequences showing the authentic location of the TG elements (N=3).

UG repeats distance from the 5'SS

Based on the extremely similar splicing pattern observed in the two constructs carrying 6 and 12 TG repeats it was hypothesized that the UG repeats might influence splicing in correlation with their proximity to the 5'SS. Therefore using the same procedure, the TG6 and TG12 repeats were inserted at 8 nucleotides from the 5'SS (Figure 9a) and the resulting constructs were again used for *in vitro* splicing assay (Figure 9b). A distance of +8 nucleotides does not perturb the 5'SS consensus sequence (from -3 to +6), but given the great proximity to the 5'SS can potentially interfere by sterical hindrance with U1snRNP recruitment. However, the splicing pattern observed with these constructs resembled the one observed with the constructs carrying TG repeats at greater distance from the 5'SS.

Similarly, no significant differences were detected between the constructs carrying different TG repeats length at different distances from the 5'SS using radio-labeled RNA as well (data not shown).

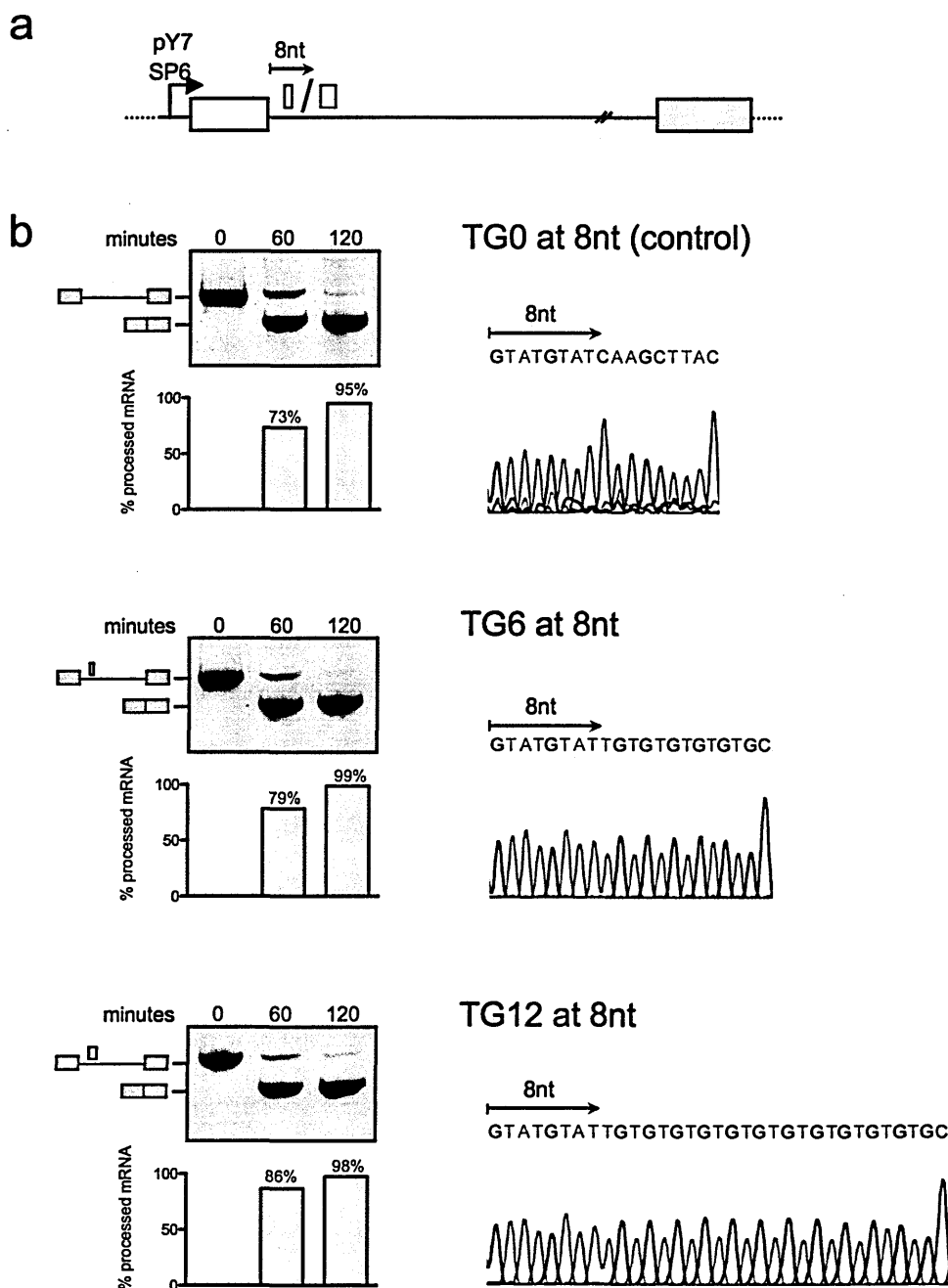


Figure 9. Insertion of TG6 and 12 at 8 nucleotides from the 5' end of exon 2 does not affect exon 2 processing. a) Schematic representation of the minigenes used. b) Splicing assay of pY7 constructs carrying TG repeats at 8 nucleotides from the 5' end of the exon 2 and the TG0 control. For each construct (TG0, TG6 and TG12), is reported the splicing pattern, the quantification of the bands and the sequences showing the authentic location of the TG elements (N=3).

RNAPII coupled transcription-splicing assay with TG0, TG6 and TG12 at 8 and 18 nucleotides form the 5'SS

In order to test whether differences in splicing regulation might be visible in conditions that closely resemble the splicing dynamics in the cells, the RNAPII coupled transcription-splicing system was used [261]. In the coupled system the DNA (not pre-transcribed RNA) is added to nuclear extract in optimized splicing conditions and in presence of alpha-32P-UTP. The DNA is then transcribed by the RNAPII, which is endogenously present in the nuclear extract, and spliced into mRNA. In this co-transcriptional splicing system the nascent RNA synthesized by RNAPII is more efficiently spliced with respect to the one transcribed by phage RNA polymerases. The enhanced efficiency is likely due to the fact that the nascent RNA is immediately and quantitatively directed into the spliceosome assembly pathway instead of being assembled into non-specific hnRNP complexes, which are inhibitory for spliceosome assembly [261].

To perform these coupled transcription-splicing assay the four pY7-based constructs with TG repeats and the TG0 control described in Figures 8a and 9a were inserted into the pcDNA3 vector because the presence of a strong promoter such as CMV is required for this method (Figure 10a).

Subsequently, equal amounts of the PCR-amplified CMV-pY7 fragments were incubated with nuclear extract for 15 minutes to generate nascent pre-mRNA, followed by addition of alpha-amanitin to block further transcription and to let splicing occur. Different time points were collected in order to track the splicing process (25-35 and 30-45 minutes). After 30 minutes the spliced mRNA was detected and by 45 minutes splicing was nearly complete. However, there were no detectable differences in the splicing pattern between the TG0 control and the four constructs carrying TG repeats (Figure 10b).

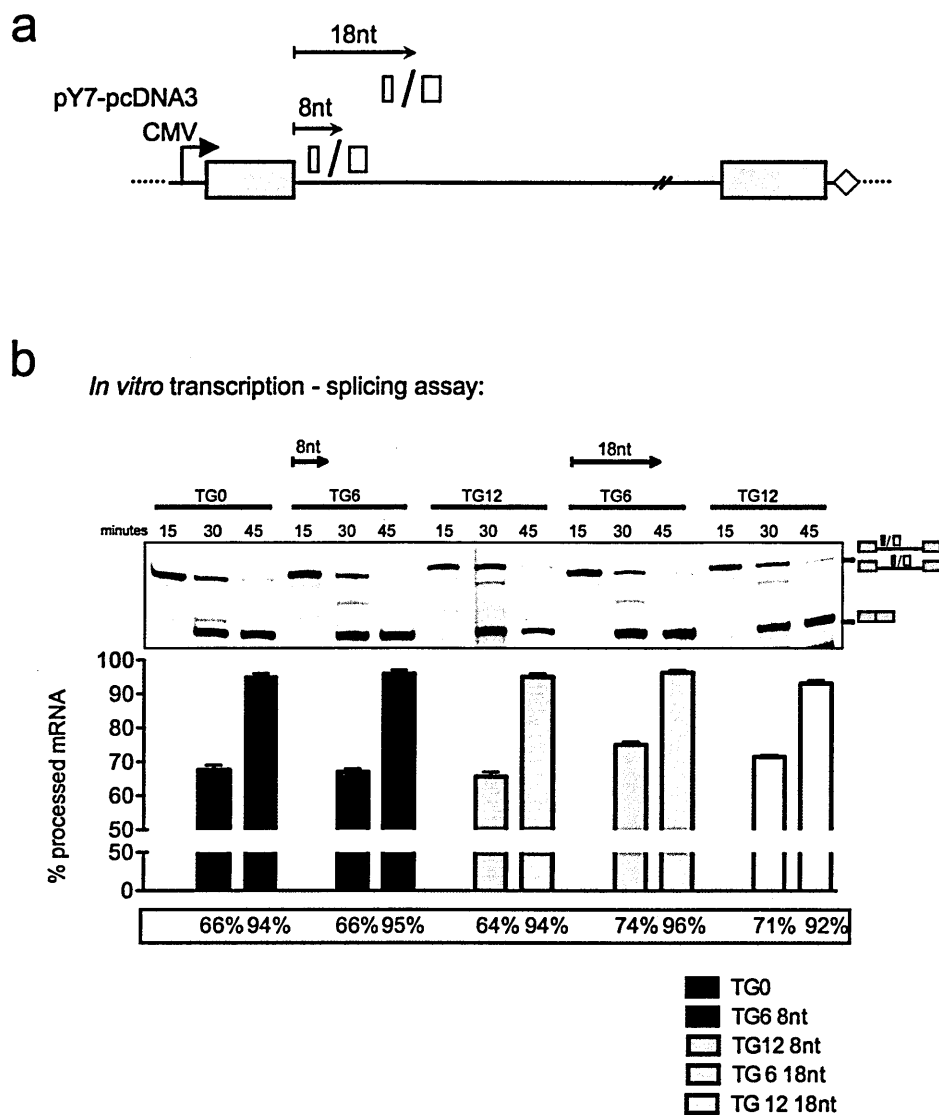


Figure 10. Insertion of TG6 and 12 at 18 and 8 nucleotides from the 5' end of exon 2 does not affect exon 2 processing. a) Schematic representation of the minigenes used. b) RNAPII coupled transcription-splicing assay of pY7 constructs carrying TG repeats at 18 and at 8 nucleotides from the 5' end of the exon and the TG0 control (N=2).

Therefore, the presented in vitro results suggest that the simple insertion of UG repeats at the 5'SS of the alpha-tropomyosin exon 2, which carries a very strong/splicing optimized 5'SS, has no significant effect on the splicing outcome of the upstream exon.

Correlation between UG repeats effect and 5'ss strength

UG repeats and 5'SS strength: selection of the BRCA1 exon 12 splicing system

The pY7 reporter minigene used for the *in vitro* experiments represented an optimized artificial system carrying an extremely well defined 5'SS consensus sequence. In fact, bioinformatic prediction of the intrinsic strength of its 5'SS reported an optimal consensus value (CV=1=100%). Therefore, this well defined 5'SS might not require the presence of additional *cis*-acting elements such as the UG repeats to be efficiently recognized by the splicing machinery.

Based on this observation, a natural substrate was used in order to check whether, in a more natural context, the presence of UG repeats in proximity of a 5'SS consensus sequence can influence the splicing outcome. The first system to be studied has been the BRCA1 exon 12, which carries two GT repeat elements in its downstream region: a GT6 at 9 nucleotides from the 5'SS followed by a GT4 at 22 nucleotides. In normal conditions exon 12 is constitutively included in the BRCA1 mRNA. Interestingly, in the BRCA1 system a natural and documented mutation was found to weaken the strength of the exon 12 5'SS in terms of consensus value.

Generally, mutations in BRCA1 gene correlate with a strong increase of breast and ovarian cancer risk. The majority of the mutations described in literature are non-sense and frame-shift mutations whereas only 4% of the identified genomic variations are splice site variants. Among these variants, the BRCA1 G→A mutation at position -1 of exon 12 (-1G>A; c.4304G>A BIC nomenclature; or c.4185G>A systematic nomenclature) identified by Gayther in 1999 was chosen [325]. The G nucleotide at position 4304, the last nucleotide of

exon 12, is conserved in 86% of splice sites in mammals. Previous published data had reported that the abolishment of this highly conserved 5'SS consensus sequence in patient's lymphoblasts leads to the partial skipping of the upstream exon where the mutation is located [326]. This -1G>A mutation has been detected in Czechs, German and American populations and is currently considered a pathogenic mutation, predisposing carriers to breast and ovarian cancer [327, 328]. In these patients, the expected (but yet to be proven) effect of the mutation at the protein level is the production of a shorter product, due to the introduction of a stop codon at position 1373 (p.Ser1363fsX10) [327].

BRCA1 exon 12 *in vitro* traditional splicing assay

For this study, both wild type (WT) and -1G>A mutated (MUT) BRCA1 exon 12 5'SS were used to produce constructs suitable for *in vitro* splicing assay. The pY7 exon 2 5'SS consensus sequence, which consist of 9 partially conserved nucleotides at the exon/intron boundary, was substituted with the one of the BRCA1 exon 12. To better mimic the natural BRCA1 context, some additional exonic and intronic regions were also changed for a total of 69 nucleotides (18 exonic nucleotides, 51 intronic nucleotides) (Figure 11).

The TG0 negative control, the WT and MUT BRCA1 5'SS constructs were then tested using *in vitro* splicing assays as previously described. Nearly identical splicing efficiency was observed for the BRCA1 WT construct and the TG0 control, which carries the splicing-optimized 5'SS consensus sequence (CV=1=100%) and no TG repeat elements (Figure 12a). Consistent with this observation, the bioinformatic prediction of the BRCA1 WT consensus sequence revealed a sub-optimal strength (CV=0.95=95%). As shown in Figure 12a, the pre-mRNA was more than 80% processed after 60 minutes and almost completely processed after 120 minutes, both in the TG0 control and in the BRCA1 WT construct. Therefore,

the presence of the TG repeats in the BRCA1 WT construct did not affect the splicing efficiency, in respect to the TG0 control (Figure 12a). This result suggest that the BRCA1 WT 5'SS consensus sequence is naturally very well defined and ensures the constitutive inclusion of exon 12 in normal conditions.

On the other hand, in the BRCA1 MUT construct the pre-mRNA was processed in a slower manner compared to the BRCA1 WT and the TG0 control constructs. In the BRCA1 WT after 60 minutes 84% of the pre-mRNA was processed, whereas only 57% of it was processed in the BRCA1 MUT and this difference was statistically significant ($p=0.0151$). Similarly, after 120 minutes 96% of the pre-mRNA was processed in the BRCA1 WT and 78% in the BRCA1 MUT constructs ($p=0.0147$) (Figure 12a). Additionally, earlier time points during splicing (20, 30 and 40 minutes) were checked and differences in splicing efficiency could be detected already at 20 minutes (Figure 12b). Therefore, these observations indicate that the -1G>A mutation weakened the 5'SS strength (predicted $CV=0.38=38\%$) causing a delay in intron processing.

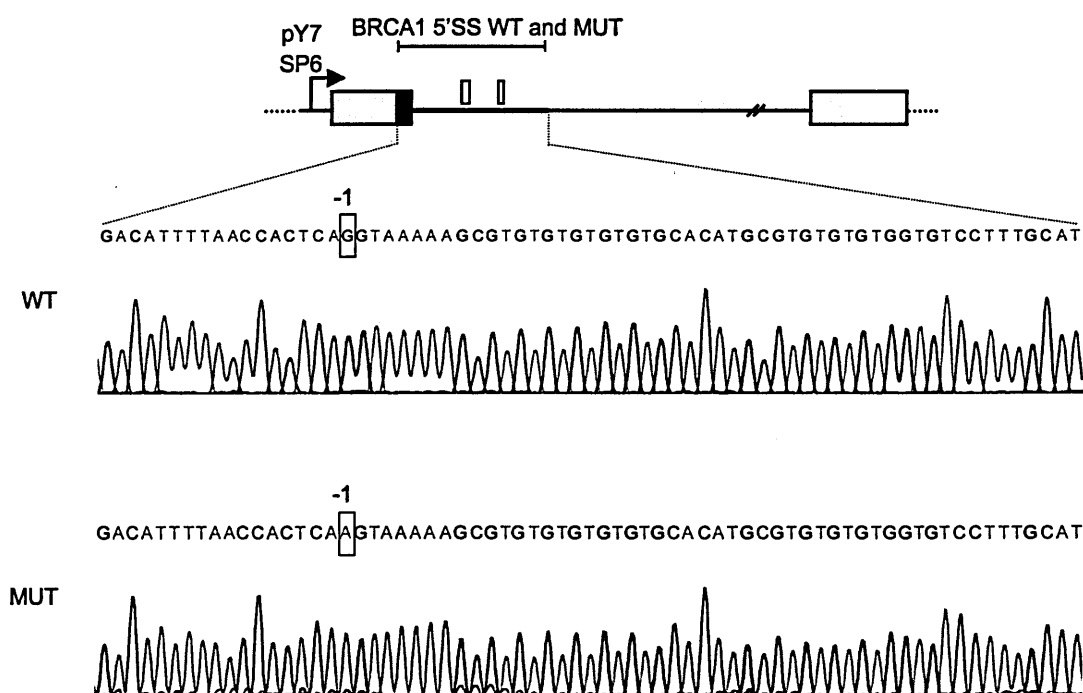
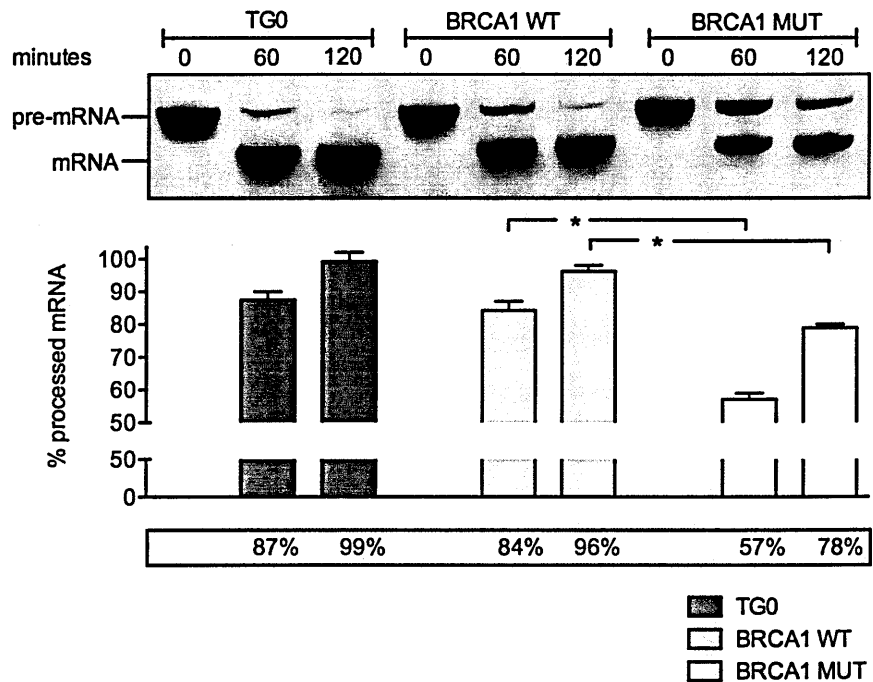


Figure 11. BRCA1 5'SS minigene constructs (WT and MUT). The sequence of the BRCA1 5'SS inserted into the pY7 reporter minigene is illustrated. The mutation G>A in position -1 and the TG repeat elements is also indicated. The GT repeats can be easily spotted within the sequence.

a



b

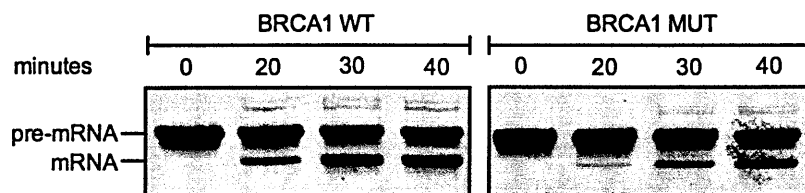
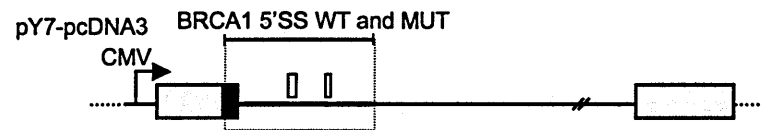


Figure 12. The mutation -1G>A weakens the BRCA1 5'SS and affects exon processing. a) Splicing assay of the two BRCA1 constructs (WT and MUT) and the pY7 control. Changing the optimized pY7 exon 2 5'SS consensus sequence with the BRCA1 WT once did not affect the correct processing of the exon. The insertion of the BRCA1 MUT weak 5'SS instead results in exon processing inhibition (N=3). b) Differences among the two constructs WT and MUT are also visible at earlier time points (N=1). The asterisks indicate statistical significance ($p<0.05$ *).

BRCA1 exon 12 *in vitro* coupled RNAPII transcription-splicing assay

In order to perform *in vitro* coupled transcription-splicing assay the pY7-based systems were inserted in the pcDNA3 vector (Figure 13a) and subsequently assessed for splicing as previously summarized (Figure 13b). After 15 minutes, both BRCA1 WT and MUT DNAs were equally transcribed by RNAPII into pre-mRNAs suggesting that the mutation does not affect proper transcription. Subsequently, splicing was allowed for 30 and 45 minutes. In the BRCA1 WT construct, splicing of the intron occurred efficiently and the pre-mRNA was almost completely spliced into mRNA after 45 minutes (Figure 13b). In the BRCA1 MUT construct splicing also occurred but in a slower manner further indicating that the presence of the mutation causes a delay in intron processing by affecting the splicing machinery (Figure 13b).

a



b

In vitro transcription - splicing assay:

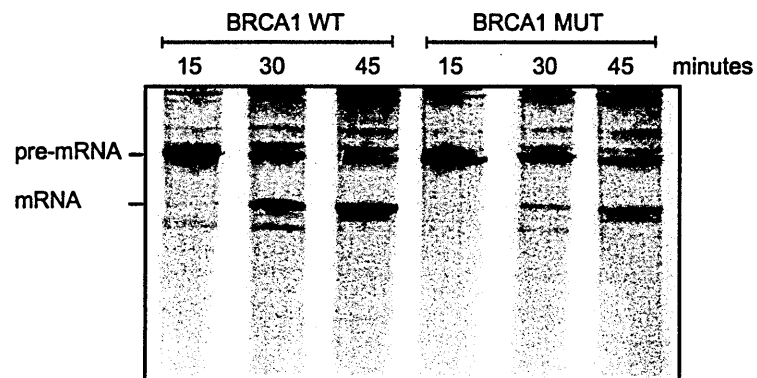


Figure 13. BRCA1 5'SS MUT presents delay in exon processing. a) Splicing representation of the minigenes used. b) Coupled splicing-transcription assay of the two BRCA1 constructs. Again a delay in exon processing is observed in the BRCA1 MUT construct with respect to the WT one (N=3).

Full length BRCA1 exon 12 *in vitro* splicing assay and *in vitro* coupled transcription-splicing assay

As previously mentioned, in patients carrying the -1G>A mutation partial skipping of the exon 12 was observed [326]. Based on this evidence, it was hypothesized that the delay in intron processing observed in the BRCA1 MUT constructs could be representative of the *in vivo* exon 12 skipping situation. To confirm this hypothesis, two pY7 constructs carrying the full length BRCA1 exon 12 (WT and MUT) were engineered in a three exons-two introns system (Figure 14a). Subsequently, these constructs were used for both traditional *in vitro* splicing and coupled transcription-splicing assays.

In the traditional *in vitro* splicing assay, by 60 minutes exon 12 was mostly included into the resulting mRNA in the BRCA1 WT construct, as expected. Also a certain amount of exon skipping was observed. On the other hand, exon 12 was completely skipped in the BRCA1 MUT construct (Figure 14b).

In the *in vitro* coupled transcription-splicing assay, both DNAs were equally transcribed into pre-mRNAs. In the BRCA1 WT construct both exon 12 inclusion and skipping were detected, whereas in the BRCA1 MUT construct exon 12 was completely skipped (Figure 14c).

Both *in vitro* approaches therefore indicate that the delay in intron processing previously observed in the two exons-one intron systems can be traduced into exon 12 skipping when using the whole BRCA1 exon 12 sequence in a three exons-two intron system. Moreover, this splicing pattern successfully mimics the aberrant splicing observed *in vivo* in patients carrying the -1G>A mutation.

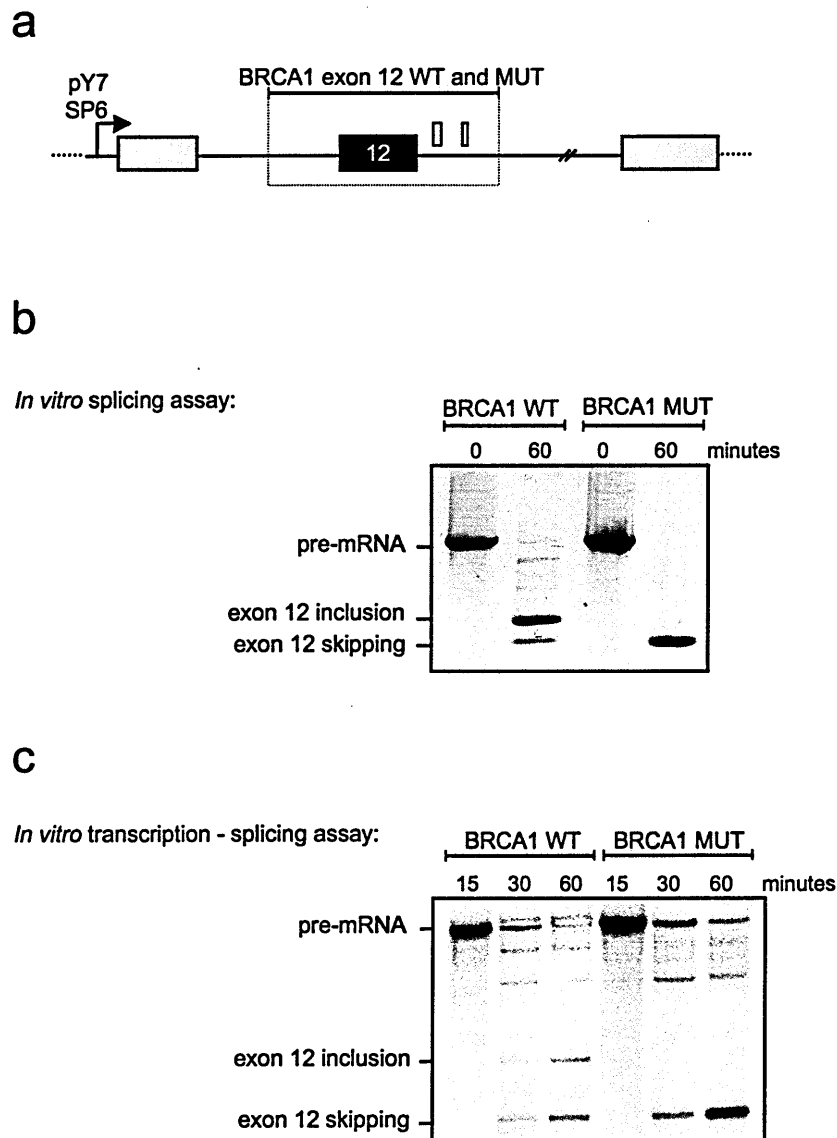


Figure 14. The mutation -1G>A leads to exon 12 skipping in constructs carrying the full length exon 12. a) Schematic representation of the minigene used (three exons-two introns). b) Splicing assay reveals exon 12 skipping in presence of the mutation (BRCA1 exon 12 MUT). A mild skipping is also observed in this case with the WT construct (N=2). c) Coupled transcription-splicing assay also reports complete exon 12 skipping in the MUT construct. Additional splicing intermediates are also visible on the gel (N=4).

The results obtained with the BRCA1 exon 12 constructs confirm that the presence of the -1G>A mutation at the 5'SS of an exon is the direct cause of aberrant splicing. Moreover, consistent with the previous reported data on TG6 and TG12 repeats at 8/19 nucleotides, the similar splicing pattern observed for the TG0 control and the BTCA1 WT construct carrying the UG repeats at the 5'SS, indicate that UG repeats located downstream of an extremely well defined 5'SS are likely not required for splicing regulation.

Nonetheless, the BRCA1 minigene constructs successfully mimicking the in vivo situation represent a suitable system for further studies on the effects of the UG repeats and of the UG repeats-binding proteins on splicing regulation, in correlation with the 5'SS consensus strength.

Effect of UG repeats-binding proteins on BRCA1 exon 12 processing

UG repeats-binding proteins affect splicing of BRCA1 exon 12

Once confirmed that both systems (with the full length exon 12 and with the 5'SS consensus sequence of exon 12, Figures 12, 13 and 14) were accurately representing the *in vivo* context, I investigated the effect of *trans*-acting factors binding to the auxiliary UG repeat elements on 5'SS definition.

For further *in vitro* studies, the two exons-one intron pY7 system carrying the BRCA1 5'SS consensus sequence was selected since the shorter length of the constructs better suits the *in vitro* experiments with nuclear extract.

The previously presented results indicate that the introduction of UG repeats in proximity of a strong and well defined 5'SS consensus sequence does not significantly affect splicing (see the artificial systems TG6 and TG12 versus the TG0 control and the BRCA1 WT versus the TG0 control, Figures 8-13). On the other hand, the -1G>A mutation was shown to be responsible for the aberrant splicing outcome observed in the BRCA1 MUT construct. Therefore, it was next investigated whether the presence of UG repeats at the 5'SS of exon might influence the splicing outcome in correlation with the 5'SS strength. In particular the UG repeats were considered as a binding site for *trans*-acting proteins.

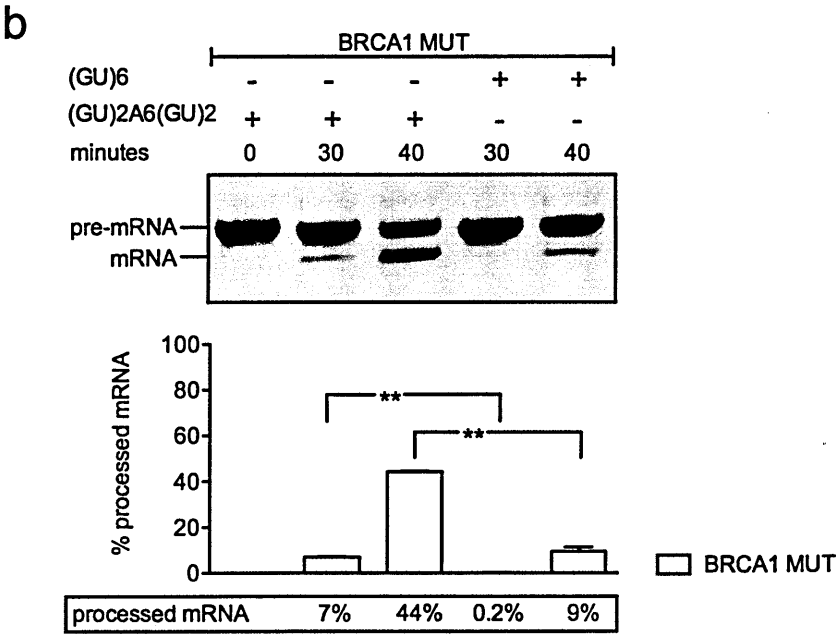
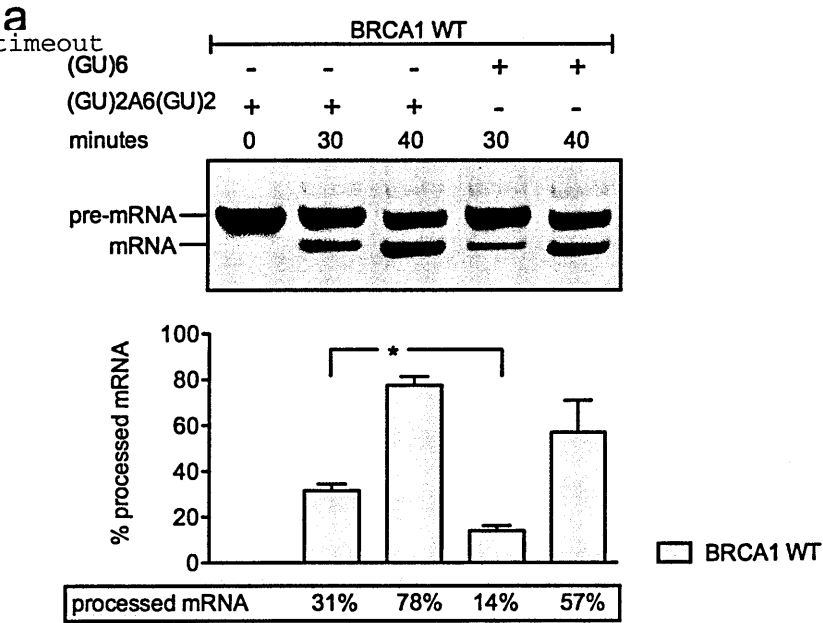
To test whether *trans*-acting factors binding the UG repeats play a role in the pre-mRNA processing events in correlation with the 5'SS strength, the nuclear extract used for splicing assay was pre-incubated with GU6 RNA oligonucleotides. The rationale of this competition assay is that the UG repeats-binding proteins present in the nuclear extract would bind to the GU6 oligonucleotides and therefore become partially unavailable for binding to the UG sequences in the

BRCA1-pY7 splicing constructs. A GU2A6GU2 oligonucleotide was used as a negative control.

After pre-incubation the nuclear extract was used in a traditional *in vitro* splicing assay. When compared to the results obtained with the GU2A6GU2 control oligo, the splicing pattern of the BRCA1 WT construct was affected by the presence of the oligonucleotides GU6 in the nuclear extract (Figure 15a). In fact, by 30 minutes 31% of the pre-mRNA was processed in the control sample but only 14% of it was processed in presence of GU6 oligonucleotides ($p=0.04$). Similarly, the processing of the intron decreased from 78% to 57% after 50 minutes (Figure 15a). This observation indicate that in this particular context the removal of UG repeats-binding proteins can exert an inhibitory effect on exon processing, even in presence of a strong 5'SS. Therefore, the presence of UG repeats at the 5'SS may favour the exon processing through the recruitment of *trans*-acting factors.

Most strikingly, the removal of UG repeats-binding proteins strongly affected splicing in presence of the weak consensus sequence of the BRCA1 MUT construct. In this system, a marked reduction of the pre-mRNA processing rate was observed. In particular, compared to the results obtained with the control oligo, pre-mRNA processing dropped from 7% to 0.2% in presence of the GU6 oligonucleotides ($p=0.0029$) after 30 minutes and from 44% to 9% after 50 minutes ($p=0.0028$) (Figure 15b). These observations highlight the importance in 5'SS definition of the UG repeats acting as *trans*-acting factors recruiters in presence of a weak consensus sequences. In fact, when the 5'SS is poorly defined or weakened by a mutation, auxiliary elements such as the UG repeats can become important elements for the upstream exon 5'SS definition.

ERROR: timeout
OFFENDING COMMAND: timeout
STACK:



UG repeats-binding proteins affect spliceosomal assembly

Based on the observation that the UG repeats can help the splicing machinery through the recruitment of UG repeats-binding proteins, I next analyzed the spliceosome assembly on both BRCA1 WT and MUT constructs, both in absence and in presence of UG repeats-binding proteins (Figure 16).

Spliceosomal constituents are recruited stepwise to the pre-mRNA substrate and remodeled multiple times in order to properly insure the final production of the spliced mRNA [329, 330]. Briefly, when the pre-mRNA is initially added to HeLa cells nuclear extract the H complex assembles on ice or at 30°C [331]. This H complex consists almost exclusively of hnRNP proteins and is considered an unspecific non-functional intermediate in the spliceosome assembly pathway. The earliest detectable functional intermediate in spliceosome assembly is the E complex, where the 5' and 3'SSs are recognized by U1 snRNP and the heterodimeric complex U2AF35/65 respectively, thus committing the pre-mRNA to the splicing pathway. Mutations of either the 5' or the 3'SSs are reported not to abolish E complex assembly [332, 333].

The E complex is rapidly converted to the A complex, characterized by the ATP-dependent recruitment of U2 snRNP to the branch point sequence. Subsequent to A complex formation, the U4/U6 and U5 snRNPs are recruited as a preassembled complex U4/U6.U5 tri-snRNP forming the B complex. A distinct role for the 5'SS in the B complex formation was highlighted in early studies where mutations of the 5'SSs resulted in accumulation of the A complex and slower formation of the B complex [334, 335]. Afterward, the spliceosome undergoes catalytic activation through major conformational and compositional rearrangements [336, 337]. The activated B complex (reported as B* complex) promotes the first catalytic step of splicing and finally the C complex leads to the formation of spliced mRNA in an ATP-mediated manner. The spliceosome complex then dissociates and the

different components are recycled for additional rounds of splicing.

In order to perform a spliceosome assembly assay, the BRCA1 WT and MUT constructs were firstly transcribed with SP6 polymerase in presence of UTP32. Secondly, a splicing reaction was carried out using bulk nuclear extract pre-incubated with either GU6 oligonucleotides or control GU2A6GU2 oligonucleotides. Three time points (0, 5 and 30 minutes) of each reaction were analyzed on a native gel in order to detect the complexes (Figure 16). In the BRCA1 WT construct, at 0 minutes the H complex was equally detected in the control and in the GU6 oligonucleotides-treated samples. After 5 minutes of incubation, complex A was equally detected as well. At 30 minutes the B complex was slightly visible both in the control and in the GU6 oligonucleotides-treated samples. The A complex was still vaguely present after 30 minutes in presence of GU6 oligonucleotides but it was not longer visible in the control.

In the BRCA1 MUT construct the H complex at 0 minutes and the A complex at 5 minutes were equally assembled. Most interestingly, at 30 minutes in the control sample the A complex was converted in B complex and only loosely detectable, but in presence of the GU6 oligonucleotides a marked accumulation of the A complex was observed (Figure 16). Interestingly, no significant difference were observed between the BRCA1 WT and MUT constructs in the negative control samples, indicating that the lone presence of the mutation did not affect the initial spliceosome assembly.

Significantly, the observation that the A complex strongly accumulates in presence of GU6 oligonucleotides, but not in presence of the control, only in the BRCA1 MUT construct carrying the weak 5'SS suggests that UG repeats-binding proteins may favour the transition between the A complex and late complexes.

Spliceosome assembly assay:

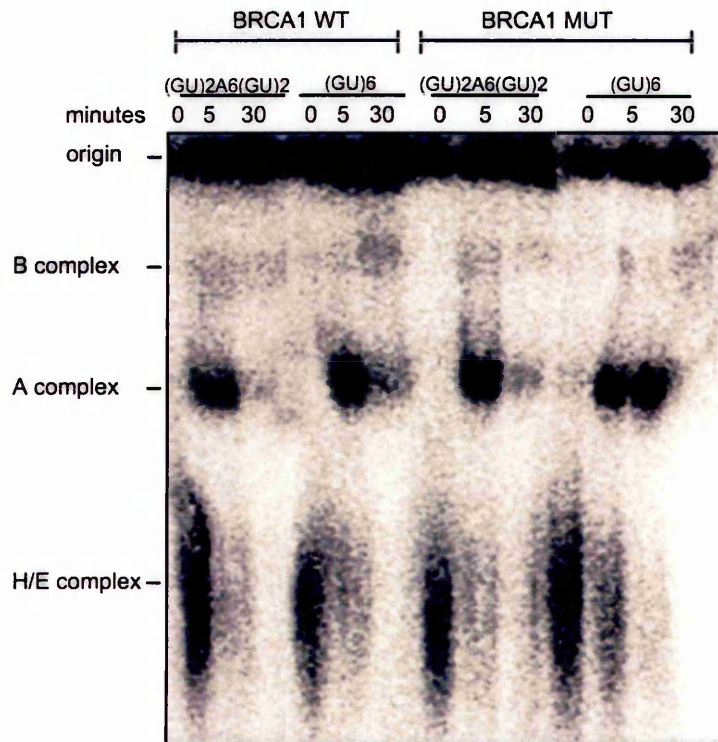


Figure 16. RNA competing oligo GU6 affects spliceosome assembly. Spliceosome assembly assay showing the effect of the competing oligo GU6 on the BRCA1 5'SS WT (a) and MUT (b) constructs. In the BRCA1 5'SS MUT spliceosome formation is stuck in A complex (N=1).

The effect of UG repeats-binding proteins removal on both splicing and spliceosome assembly strongly suggests a role for the UG repeats-binding proteins in enhancing the processing of the upstream exon. In addition, the effect of the UG repeat-binding protein was solidly related to the 5'SS strength indicating that auxiliary elements such as the UG repeats help 5'SS definition in poorly defined exons.

Effect of TDP-43 levels on BRCA1 exon 12 splicing regulation

Selection of the most suitable BRCA1 splicing system

As reported in the introduction, proteins binding to the UG/TG repeats identified so far include TDP-43, CELF1, CELF2, PTB, SRp55 and FUBP3. All these proteins have various effects (i.e. act as splicing silencer or enhancers) and might compete with each other for their preferential binding site to ultimately affect the splicing outcome. Among these UG repeats-binding proteins this study focused on TDP-43, to better define its potential involvement in 5'SS definition and splicing regulation.

The previously reported *in vitro* results using competitor GU6 RNA oligonucleotides presented the limitation of potentially affecting all the variety of UG/TG repeats-binding proteins present in the extract and not only TDP-43. Therefore, in order to selectively observe TDP-43 effects on splicing regulation of the BRCA1 constructs, *in vitro* experiments in siRNA treated cell cultures were performed. In Baralle's laboratory a common procedure based on transfection of pTB-based minigenes in HeLa cells upon TDP-43 depletion/overexpression has been successfully used over the years to study TDP-43-mediated splicing regulation in different systems of interest [88, 203, 272]. Therefore, this same procedure was applied to the study of TDP-43 involvement in BRCA1 exon 12 processing.

As for the *in vitro* with nuclear extract systems, it was first established the most suitable system capable of mimicking the *in vivo* context to use for experiments in cell cultures. To do so, different splicing systems for the BRCA1 exon 12, based on pY7, pcDNA3 and pTB vectors were tested. Subsequently, all the constructs were transiently transfected in HeLa cells. After 24 hours, total RNA

was harvested, retro-transcribed and PCR-amplified; finally the splicing products were separated on 2% agarose gels. In some cases, TDP-43 was silenced through RNA interfering technique previous construct transfection in order to detect the potential TDP-43 involvement in exon processing. A control-siRNA was always used as a negative control.

The BRCA1 WT and MUT pY7-based splicing systems inserted in the pcDNA3 vector under the control of the CMV promoter (constructs formerly used for coupled transcription-splicing system) were transfected in HeLa cells in presence of the control siRNA (siLuciferase) or the siTDP-43 (Figure 17a). In this splicing system both BRCA1 WT and MUT constructs were efficiently processed in control siLuciferase-treated cells. Also in this particular context, the mutation in the BRCA MUT construct did not seem to affect the splicing outcome as previously observed in control condition. In the TDP-43-silenced cells a reduced amount of PCR product was observed for the BRCA1 MUT construct. Since the performed RT-PCR does not represent a quantitative technique, real time PCR experiments were also performed but did not report significant differences among the different samples.

No differences in the splicing pattern of the BRCA1 WT and MUT were observed by inserting the two pY7-based splicing systems into the pTB vector (Figure 17b) and also by inserting it into an exonic cassette within the pTB vector (Figure 17c).

These results indicated that the pY7-based splicing-optimized systems might not be suitable for cell culture experiments. In fact, no cases of pY7-based systems have ever been reported in literature as a successful tool in cell cultures experiments. Nevertheless, the use of different vectors has already been reported to affect the splicing outcome of the studied systems, and also to produce conflicting results (i.e. [182] and [183]). In fact, the structure of different minigene

systems might affect in unpredictable ways the splicing outcome. These observations highlight the importance of evaluating the splicing outcome of the system of interest in more than one reporter minigene in order to select the one that better mimics the *in vivo* context.

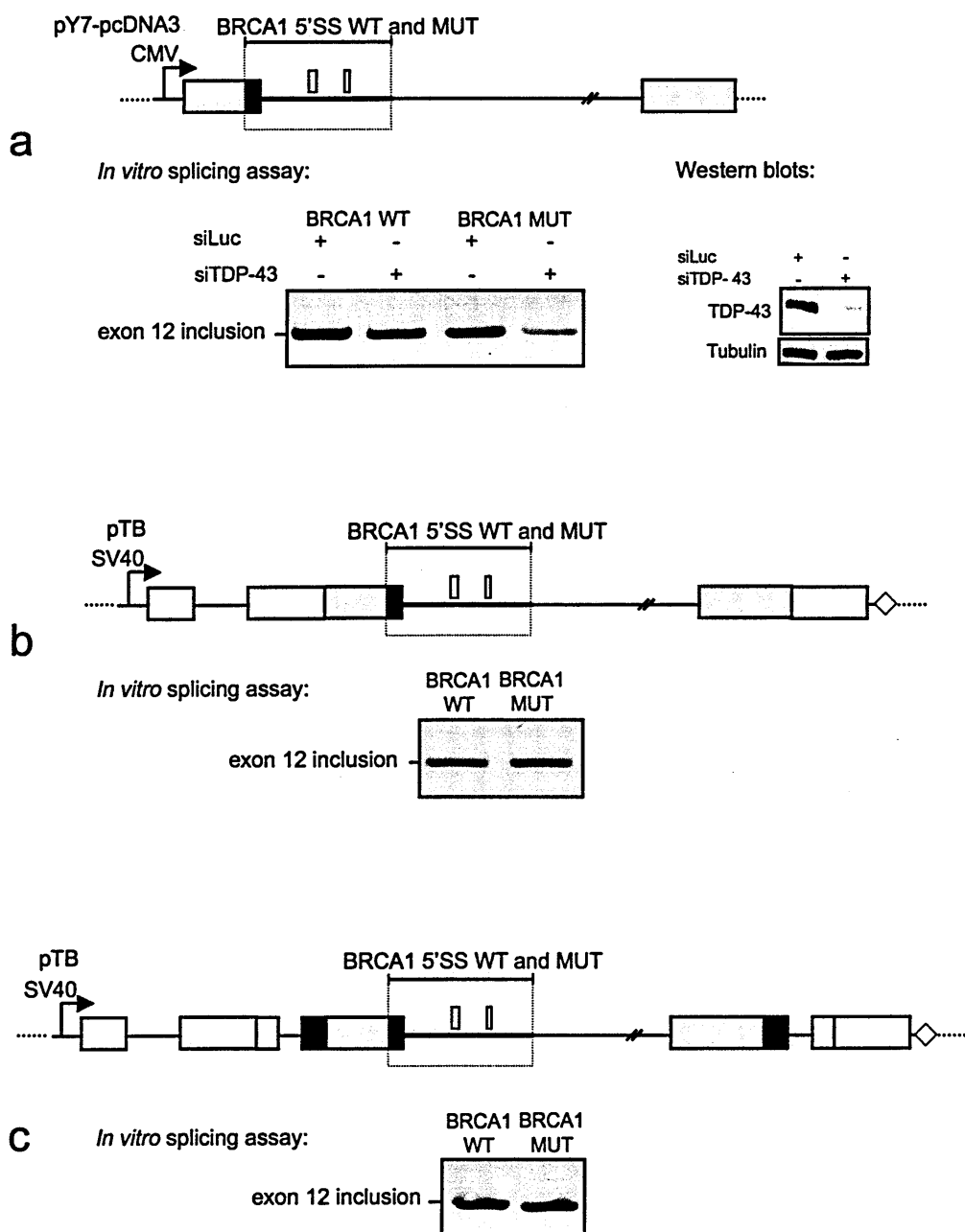


Figure 17. Splicing analysis of BRCA1 5'SS WT and MUT in different reporter minigene systems. a) A schematic representation of the pY7 minigenes inserted into the pcDNA3 vector, the splicing pattern and the western blots detecting TDP-43 silencing. b) A schematic representation of the BRCA1 5'SS WT and MUT inserted into the pTB vector and the reported splicing pattern in HeLa cells. c) A schematic representation of the BRCA1 5'SS WT and MUT inserted into an exonic cassette within the pTB vector and the splicing assay. The tested constructs do not represent a suitable tool for further studies (N=1).

Therefore in further studies, the two exons-one intron pY7 system was abandoned in favour of an extended genomic context of the BRCA1 exon 12 in a three exons-two intron system. These BRCA1 exon 12 cassettes (WT and MUT) with the flanking alpha-tropomyosin exon 2 and 3 were inserted in the pcDNA3 (pY7-based construct used for the coupled transcription-splicing system as well). Additionally, the same BRCA1 cassettes were inserted in the pTB vectors, which carries the alpha-globin/fibronectin flanking exons (Figure 18a, 18b).

In the pcDNA3 context the BRCA1 exon 12 WT was, as expected, equally processed both in the control and in the siTDP-43 treated samples, thus further confirming the previous results. The BRCA1 exon 12 MUT was not recognized by the spliceosomal complex and therefore skipped in the final mRNA both in the control and in the TDP-43 silenced experimental contexts. These observations indicate that the mutation itself directly causes exon 12 skipping and additionally that the silencing of TDP-43 does not enhance or inhibit exon 12 recognition in this context (Figure 18a).

In the pTB context the BRCA1 exon 12 WT was again properly recognized and correctly processed. On the other hand, the BRCA1 MUT presented both regular exon 12 inclusion (57%) and exon 12 skipping (Figure 18b), consistent with previous observations (De Conti, personal communication). Additionally, a cryptic 5'SS located 58 nucleotides downstream to the authentic one in intron 12-13 was also recognized (CV=0.27=27%). This cryptic 5'SS corresponds to nucleotide 37848 (GenBank accession number L78833).

Silencing of TDP-43 did not affect the BRCA1 WT construct but resulted in the complete abolishment of exon 12 recognition, leading to exon 12 skipping in the BRCA1 MUT construct (Figure 18b).

Among the ten minigene systems tested in cell cultures (Figures 17a-c, 18a-b), the splicing pattern obtained with the full length exon 12 inserted in the pTB vector was the one better mimicking the *in vivo* pattern detected in patients (Figure 18b). In fact, in patient's lymphoblasts both exon 12 inclusion and skipping were observed [326]. Based on the great similarity with the *in vivo* condition the pTB-based splicing system was selected for further studies.

Noteworthy, Claes et al. detected an in-frame insertion of 66 bp from intron 13-14 both in the control and in the patient (Figure 18c). This DNA cassette, previously referred to as pseudoexon 13a [338, 339], corresponds to nucleotides 49332-49397 (GenBank accession number L78833) and does not correspond to the cryptic splice donor site observed in Figure 18b in the BRCA1 MUT lanes. Analysis of the literature did not reveal any reported cases of 37848 cryptic recognition *in vivo*. Based on the null biological relevance of this cryptic site and on the fact that its recognition did not occur in any of the other minigenes tested in cells culture nor *in vitro*, the corresponding splicing product was not considered in further studies under the assumption that the 37848 cryptic activation was likely driven by the pTB vector structure itself.

Effects of TDP-43 on BRCA1 MUT exon 12 processing

In the selected pTB-based BRCA1 exon 12 MUT construct, TDP-43 silencing abolished exon 12 recognition, indicating that in this splicing system the protein exerts an enhancing effect on exon 12 recognition.

To further examine TDP-43 involvement in BRCA1 exon 12 definition and to better define which TDP-43 domain is directly involved in splicing regulation, flag-tagged and siRNA-resistant TDP-43 constructs were overexpressed in cell culture experiments. Four different recombinant TDP-43 were used: i) TDP-43 wild type (WT); ii) TDP-43 with domain RRM1 and RRM2 mutated, thus deprived of its RNA-binding activity (RRM1,2M); iii) TDP-43 lacking the entire C-terminal domain (Δ C); iv) TDP-43 lacking the hnRNP binding site and the Q/N rich region (Δ 321-366).

Firstly TDP-43 was silenced and secondly the recombinant siRNA resistant TDP-43 constructs transfected together with the BRCA1 minigenes. After 24 hours total RNA was collected, retro-transcribed and PCR-amplified. Splicing products were finally separated on 2% agarose gels (Figure 19b).

Overexpression of WT TDP-43 failed to completely restore exon 12 inclusion but a partial rescue (25%) was observed further indicating that TDP-43 helps the recognition of exon 12 (Figure 19b). As expected, overexpression of RRM1,2M TDP-43 failed to recover exon 12 recognition from the TDP-43-silenced condition, indicating that the integrity of the RNA-binding domain is an absolute requirement for TDP-43-mediated splicing regulation. On the other hand, overexpression of both Δ C TDP-43 and Δ 321-366 TDP-43 partially restored exon 12 inclusion. In particular, overexpression of Δ 321-366 TDP-43 was found to better rescue exon 12 inclusion (17%) with respect to the Δ C TDP-43 (9%). Therefore, the TDP-43-mediated enhancement of exon 12 inclusion might not be

driven just by the 321-366 specific region but by a wider region in the C-terminal domain.

The western blots reported in Figure 19c show that all the four flag-tagged siRNA-resistant proteins were expressed in similar quantities in the transfected cells and that TDP-43 was efficiently knocked down. Also, the loading control protein (tubulin) was present at similar levels in control and silenced/overexpressed cells.

As an additional control, the same pattern of experiments was performed using the BRCA1 WT construct. Indeed, neither silencing nor overexpression of the four different recombinant TDP-43 proteins were able to affect exon 12 regular processing (Figure 20b and 20C for the western blot controls) in line with the previously reported results.

In summary, the experiments in cell culture indicate that UG repeats located in the downstream region of the BRCA1 MUT exon 12 can help the 5'SS definition of the poorly defined exon 12 through the recruitment of TDP-43, which acts as a splicing enhancer. In particular, TDP-43 RRM1 and RRM2 domains are required for the protein to modulate splicing. Moreover, not only the 321-366 region involved in the interaction with others hnRNP family members, but the entire TDP-43 C-terminal domain plays a role in exon 12 splicing regulation.

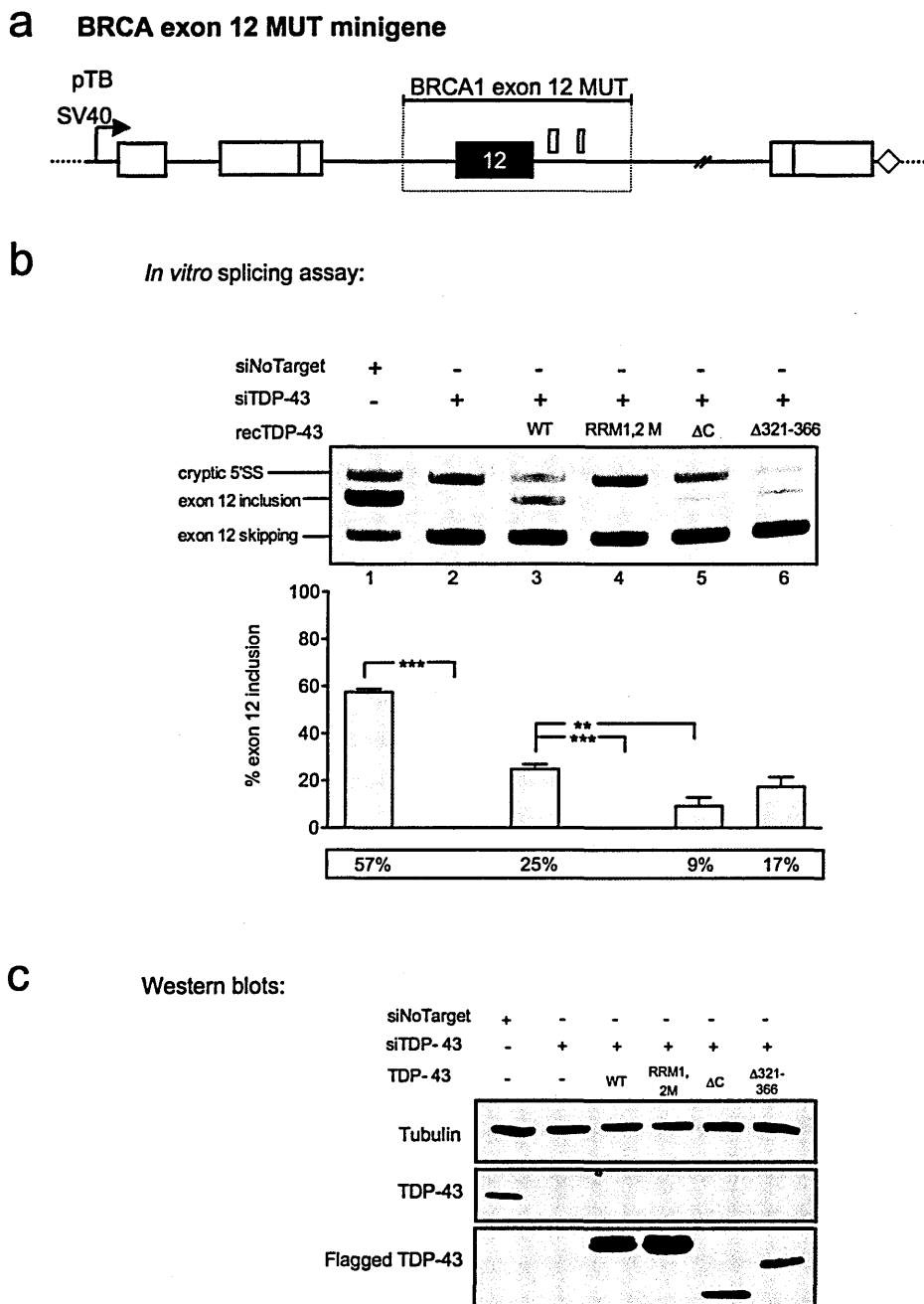


Figure 19. TDP-43 favors BRCA1 exon 12 MUT inclusion in the final mRNA product. a) A schematic representation of the BRCA1 exon 12 MUT construct in the pTB reporter minigene used for splicing assay in HeLa cells. b) Splicing assay in cell cultures reveals that TDP-43 favors exon 12 MUT inclusion. Overexpression of the protein restores, but not completely, exon 12 inclusion. c) Western blots detecting the TDP-43 silencing and flagged-tagged recombinant TDP-43 proteins overexpression (N=6). The reported results showed statistical significance ($p < 0.001$ ***).

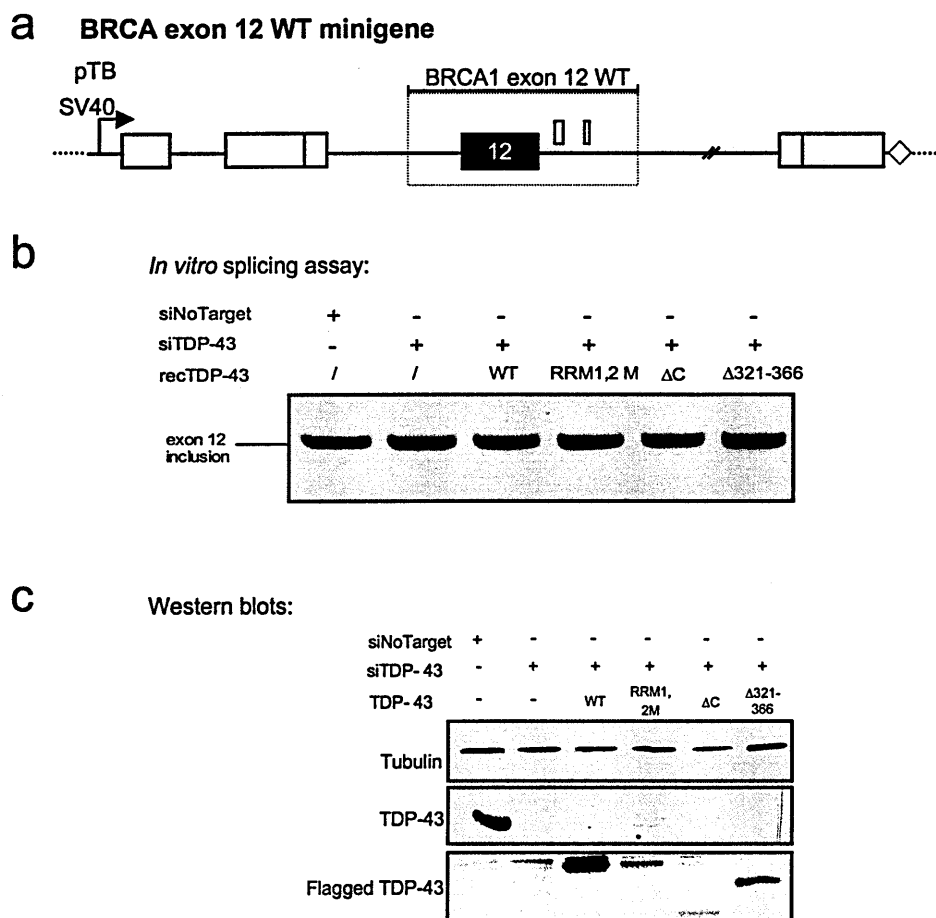


Figure 20. TDP-43 does not affect BRCA1 exon 12 WT inclusion. a) A schematic representation of the BRCA1 exon 12 WT construct in the pTB reporter minigene used for splicing assay in HeLa cells. b) Splicing assay in cell cultures reveals that TDP-43 does not affect exon 12 processing in presence of the strong 5'SS. c) Western blots detecting the TDP-43 silencing and flagged-tagged recombinant TDP-43 proteins overexpression (N=3).

Identification of new TDP-43 potential RNA targets

Identification of exons carrying UG repeats in proximity of their 5'SS through bioinformatic analysis

Our previous results have shown that the BRCA1 WT exon 12 has remained unaffected by both TDP-43 silencing and overexpression. On the other hand TDP-43, through the binding of UG repeats located near the 5'SS, enhanced the recognition of the upstream exon 12 in the BRCA1 MUT pathological context. To identify new potential TDP-43 regulated exons carrying UG elements in their 5'SS proximal intronic sequence, I performed a genome-wide bioinformatic analysis. Specifically, using a database of human genes that undergo alternative splicing of their pre-mRNAs (AltSplice database [323, 324]), 130 exons were identified carrying at least six UG/TG repeats either at the 3' or 5'SS. Among them, the 45 exons carrying UG/TG repeats at their 5'SS are reported in Table 3.

In Table 3, for each exon involved (third column), the number of UG/TG repeats in its downstream sequence is reported (fourth column). Noteworthy, in this study a downstream sequence of 70 nucleotides has been considered. Therefore, the reported number of UG/TG repeats in Table 3 refers to this interval, although in different cases the UG/TG elements extended beyond it. In addition, given that the 5'SS consensus sequence consists of 9 nucleotides spanning from position -3 to +6, the location definition (fifth column) indicates whether the UG/TG repeats are located within the consensus (within +1 and +6) or in the downstream sequence (for distances longer than +6). Finally, the 5'SS strength in terms of consensus value (CV) was calculated using a bioinformatic tool (seventh column). Clearly, the BRCA1 exon 12 previously analyzed in this dissertation was also reported in the bioinformatics analysis.

Among the new potential TDP-43 targets identified with the bioinformatic analysis, I selected five representative exons, which were also reported as TDP-43 clusters on CLIP analysis (Ule, personal communication) and affected by TDP-43 silencing in a previous microarray analysis from Baralle's laboratory [199]. The selected exons (in bold in Table 3) represent a variable pool in terms of CVs (ranging from CV=0=0% to CV=0.97=97%), location (ranging from +4 to +34) and number of UG repeats. This significant variability allowed investigation of the potential effect of TDP-43 in different genomic contexts.

Table 3. Exons carrying UG/TG repeats at their 5'SS. The gene symbols (first column) and gene names (second column) are reported for each involved exon (third column). In the forth column the number (n) of UG/TG repeats is listed; the “_” indicates that the repeat elements are interrupted by non-repeated sequences. The fifth column reports the number of the intronic nucleotides from the 5' exon end. Given that the 5'SS consensus sequence is defined by 9 nucleotides, the location definition (sixth column) identifies whether the TG elements are in the consensus sequence (from +1 to +6) or in the proximal downstream region (from +7 to +70). Finally, the last column reports the consensus value for each 5'SS (seventh column).

Gene Symbol	Gene Name	Exon Involved	(TG)n	Location	Location Definition	CV
ACSS2	Acyl-CoA synthetase short-chain family member 2	exon 8	21	+5	5'SS	0.4
AMHR2	Anti-Mullerian hormone receptor, type II	exon 6	7_3	+21	down stream	1
ATF6B	Activating transcription factor 6 beta	exon 5	2_6	+46	down stream	0.95
BRCA1	Breast cancer 1, early onset	exon 12	6_4	+10	down stream	0.95
CDC25A	Cell division cycle 25 homolog A (S. pombe)	exon 7	8_2	+6	down stream	0.88
COL4A5	Collagen, type IV, alpha 5	exon 15	14_7	+44	down stream	0.99
COL4A6	Collagen, type IV, alpha 6	exon 22	3_13	+35	down stream	1
CST3	Cystatin C	exon 2	2_11	+3	5'SS	0.99
CST5	Cystatin D	exon 2	2_2_3_6	+3	5'SS	1
CTH	Cystathionase (cystathionine gamma-lyase)	exon 11	14	+14	down stream	0.96
DDR1	Discoidin domain receptor tyrosine kinase 1	exon 8	9	+51	down stream	0
DMC1	DMC1 dosage suppressor of mck1 homolog	exon 13	25	+18	down stream	0.94
EPHX1	Epoxide hydrolase 1, microsomal (xenobiotic)	exon 5	19	+21	down stream	0.97
ETF1	Eukaryotic translation termination factor 1	exon 7	2_14	+27	down stream	0.23

Gene Symbol	Gene Name	Exon Involved	(TG)n	Location	Location Definition	CV
GAGE8	Cancer/testis antigen 4.7	exon 3	3_5_6	+31	down stream	0.99
HLA-DRB1	Major histocompatibility complex, class II, DR beta 1	exon 2	2_9	+32	down stream	0.99
HLA-DRB3	Major histocompatibility complex, class II, DR beta 3	exon 2	2_6	+32	down stream	0.99
HMG14	High-mobility group nucleosome binding domain 1	exon 5	14	+8	down stream	0.51
HMGN1	High-mobility group nucleosome binding domain 1	exon 5	19	+8	down stream	0.51
IGFBP7	Insulin-like growth factor binding protein 7	exon 2	15	+29	down stream	0.9
JAG1	Jagged 1	exon 26	2_14	+5	5'SS	0.99
KLF8	Kruppel-like factor 8	exon 4	2_11	+30	down stream	0.98
MAP3K19	Mitogen-activated protein kinase kinase kinase 19	exon 3	9_2	+3	5'SS	0.59
MICALL1	MICAL+like 1	exon 2	2_11_12	+35	down stream	0.99
MSH6	MutS homolog 6 (E. coli)	exon 2	5_11	+19	down stream	0.99
MYH7	Myosin, heavy chain 7, cardiac muscle, beta	exon 32	6	+1	5'SS	0.89
MYL2	Myosin, light chain 2, regulatory, cardiac, slow	exon 4	8_3_2	+9	down stream	1
N4BP2L2	NEDD4 binding protein 2-like 2	exon 3	9	+7	down stream	0.86
NTRK1	Neurotrophic tyrosine kinase, receptor, type 1	exon 12	16	+34	down stream	0
ORC5L	Origin recognition complex, subunit 5	exon 6	11_2	+32	down stream	0.81
PLP2	Proteolipid protein 2	exon 4	7_14	+5	5'SS	1
PPIA	Peptidylprolyl isomerase A (cyclophilin A)	exon 2	2_7	+38	down stream	0.86

Gene Symbol	Gene Name	Exon Involved	(TG)n	Location	Location Definition	CV
PTPRU	Protein tyrosine phosphatase, receptor type, U	exon 24	2_7_5	+16	down stream	0
RBCK1	RanBP-type and C3HC4-type zinc finger containing 1	exon 5 (AS)	8	+53	down stream	0.45
RGL2	Ral guanine nucleotide dissociation stimulator+like 2	exon 11	6	+8	down stream	0.78
RXRG	Retinoid X receptor, gamma	exon 7	3_14	+25	down stream	0.15
SCAP2	Src kinase associated phosphoprotein 2	exon 9	8_6	+39	down stream	0.76
SFTPC	Surfactant protein C	exon 4	6_18	+1	5'SS	0.95
SGK2	Serum/glucocorticoid regulated kinase 2	exon 8	18_2	+4	5'SS	0.67
SIGLEC1	Sialic acid binding Ig-like lectin 1, sialoadhesin	exon 12	20	+5	5'SS	0.98
ST5	Suppression of tumorigenicity 5	exon 13	9	+51	down stream	0.99
SYF2	SYF2 homolog, RNA splicing factor (S. cerevisiae)	exon 5	2_8	+26	down stream	0.99
T3 delta	CD3d molecule, delta	exon 3 (AS)	2_18	+26	down stream	0.5
TBX19	T-box 19	exon 3	20	+7	down stream	0.98
TG	Thyroglobulin	exon 10	13_3	+27	down stream	0.54

TDP-43 in splicing regulation of five new potential targets

Five potential target exons analyzed in minigene systems

The five selected exons (EPHX1 exon 5, ETF1 exon 7, NTRK1 exon 12, RXRG exon 7 and SGK2 exon 8) together with 300 nucleotides-flanking intronic sequences were firstly amplified from the genomic DNA (from HeLa cells) and secondly inserted in the pTB vector. The pTB vector was selected for studies in cell culture. Each of the five minigene systems was transiently transfected in control cells and in TDP-43 silenced cells. Additionally, overexpression the four different recombinant forms TDP-43 was performed (WT; RRM1,2M; Δ C; and Δ 321-366 TDP-43).

In this set of experiments the main goal was to determine whether in these five new splicing systems, the UG repeats act as enhancers favoring exon inclusion in a TDP-43-mediated manner as observed in the BRCA1 MUT system, or conversely act as silencers promoting exon skipping through the recruitment of TDP-43, as reported in literature for the CFTR exon 9 and APOAII exon 3 systems. In addition, the involvement of the different TDP-43 domains in splicing regulation was also investigated.

TDP-43 overexpression induces RXRG exon 7 skipping

The RXRG (retinoid X receptor gamma) is a member of the nuclear retinoid X receptor family (RXR) together with RXRA (alpha) and RXRB (beta). The three members function as ligand-activated transcription factors and contain three conserved domains: i) a DNA-binding domain; ii) a ligand-binding domain; and iii) a transcription-activator domain. The RXRs can form either homodimers or heterodimers with many other nuclear receptors [340]. In particular, heterodimers

of RXR and retinoic acid receptors (RAR) bind to the ligand retinoic acid (RA), a metabolic product of vitamin A known to modulate gene expression of more than 500 RA-responsive genes. Upon activation by the RA ligand, the heterodimeric pair RXR:RAR binds to DNA sequences and induces or represses gene expression [341]. Concerning RXRG, three different isoforms of the protein corresponding to alternative splicing events have been observed (ENST00000465764); exon 7 is normally included in the final transcript (Figure 21a). As reported in Table 3, the downstream sequence of RXRG exon 7 presents a UG repeat element composed of a UG3 at position +25 followed by a UG14 at +36. This UG element is mostly conserved among primates, but poorly conserved among amniotes vertebrates (Figure 21b). Although the bioinformatic examination reported a very weak 5'SS ($CV=0.15=15\%$), exon 7 is normally included in the final RXRG mRNA probably because of significant ESE and ISE sequences located in the nearby sequences that compensate for the weak 5'SS consensus sequence (Figure 29 in appendix).

In order to determine the biological impact of the UG repeats and TDP-43 in this splicing system a minigene system carrying both RXRG exon 7 and 300 nucleotides of the intronic flanking region was engineered (Figure 22a) and then transfected in control, TDP-43-silenced and recombinant TDP-43-overexpressed cells (Figure 22b).

In the control sample RXRG exon 7 was fully included, resembling the wild type conditions. Silencing of TDP-43 did not seem to affect the processing of the exon that was included in the final mRNA as in the control. Based on these observations it can be excluded that TDP-43 exerts an enhancing effect on exon 7 inclusion since no exon 7 skipping was observed upon TDP-43 silencing. On the other hand, given the already complete inclusion of exon 7 in the control sample it was not possible to detect an increase in exon 7 inclusion upon TDP-43 silencing.

For this reason, it cannot to be excluded that TDP-43 might exert an inhibitory effect on the exon 7 recognition.

Most interestingly, however, overexpression of WT TDP-43 resulted in exon 7 skipping (27%). Moreover, mild exon 7 skipping was also observed by overexpressing the ΔC and $\Delta 321-366$ TDP-43, 2% and 7% respectively. These results suggest that in overexpression conditions TDP-43 can act (directly or indirectly) as a splicing silencer or alternatively that overexpressed TDP-43 can competes for the binding site with other UG repeats-binding proteins possibly involved in promoting exon 7 inclusion.

Moreover, the overexpression of RRM1,2M TDP-43 mimicked TDP-43-silencend conditions, suggesting that, as for the BRCA1 MUT, the RNA-binding function of TDP-43 is fundamental to affect TDP-43-mediated exon 7 processing (Figure 22b).

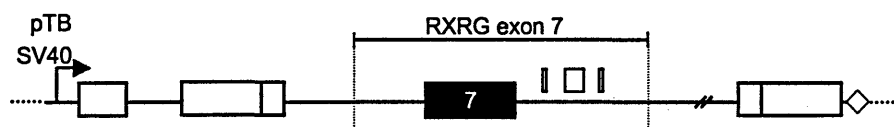
Diagram illustrating alternative splicing of exon 7. The top part shows a pre-mRNA sequence with exons 5, 6, 7, 8, and 9. Exon 7 is flanked by a 5' splice site (dotted line) and a 3' splice site (dotted line). A solid line connects the 5' splice site to the 3' splice site, bypassing exon 7, representing a constitutive splice. A dashed line connects the 5' splice site to the 3' splice site, including exon 7, representing an alternative splice. The bottom part shows the resulting mRNA sequence: 5' ... 5 6 7 8 9 3'. The text "exon 7 is constitutively included" is written below the sequence.

RXRG

Human (<i>Homo sapiens</i>):	G T A T T T G T G T G T A T T A T G T G T G T G T G T G T G T G T G T G T A T A C A
Chimpanzee (<i>Pan troglodytes</i>):	G T A T T T G T G T G T A T T A T G T G T G T G T G T G T G T G T G T G T A T A C A
Gorilla (<i>Gorilla gorilla</i>):	G T A T T T G T G T G T A T T A T G T G T G T G T G T G T G T G T G T G T A T A C A
Orangutan (<i>Pongo pygmaeus</i>):	G T A T T T G T G T G T A T T A T G T G T G T G T G T G T G T G T G T G T A T A C A
Macaque (<i>Macaca mulatta</i>):	G T A T T T G T G T G T A T T A T G T G T G T G T G T G T G T G T G T G T A T A C A
Marmoset (<i>Callithrix jacchus</i>):	G T A T T T G T G T G T A T T A T G T A T G T A T G T G T G T G T G T G T A T A C A
Chicken (<i>Gallus gallus</i>):	G C - T T A T C A - - - - -
Turkey (<i>Meleagris gallopavo</i>):	G G - T T - G T C A - - - - -
Zebra Finch (<i>Taeniopygia guttata</i>):	G G - T T - G C C A - - - - -
Mouse (<i>Mus musculus</i>):	G G A T T T A C A C A C A G T G - - - - - T G C A T A T G C A
Rat (<i>Rattus norvegicus</i>):	G G A T T T A C A C C C A C T G - - - - - T G C A C A T G C A
Rabbit (<i>Oryctolagus cuniculus</i>):	G T A T C C C T C T C T G - - - - - T G T G T A T A T A C A
Horse (<i>Equus caballus</i>):	G T G C T T G T G T G T G - - - - - G T A T G T G T A T A C A
Dog (<i>Canis familiaris</i>):	G T G C T C G T G T G T G T G T - - - - - G T G T G T G T G T C T A C A
Pig (<i>Sus scrofa</i>):	G T G T T C G T G T G T G T - - - - - T G T G T G T G T G T G T G
Cow (<i>Bos taurus</i>):	G C A - - - - T G T G - - - - - T G T G T G T G T G T G T G

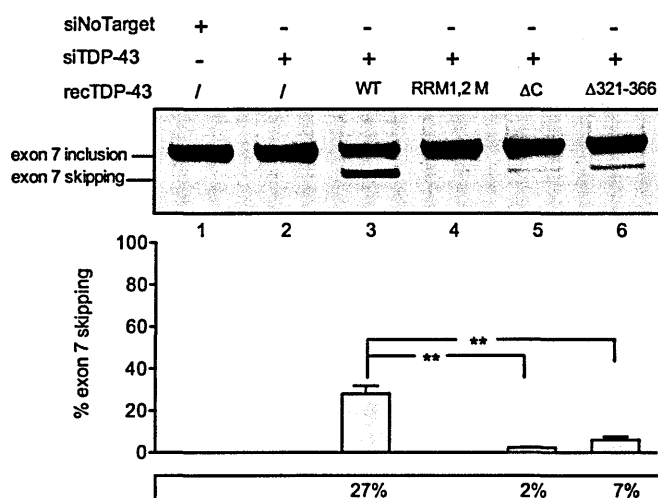
Figure 21. RXRG exon 7 constitutive splicing and UG element conservation among mammals. a) RXRG exon 7 is constitutively included in the resulting mRNAs. b) The UG element at the 5'SS is mostly conserved among primates, but poorly conserved among amniotes vertebrates.

a RXRG exon 7 minigene



b

In vitro splicing assay:



c

Western blots:

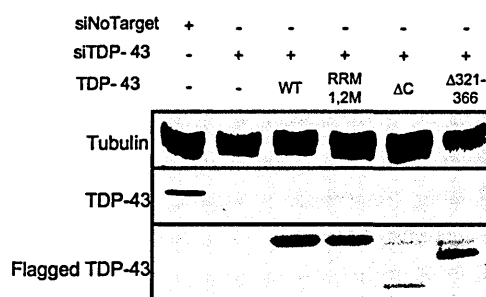


Figure 22. TDP-43 overexpression inhibits RXRG exon 7 inclusion. a) A schematic representation of the RXRG exon 7 construct in the pTB reporter minigene used for splicing assay in HeLa cells. b) Splicing assay in cell cultures reveals that, whereas TDP-43 silencing does not seem to affect exon inclusion, TDP-43 overexpression enhances exon 7 skipping (N=2). c) Western blots detecting the TDP-43 silencing and flagged-tagged recombinant TDP-43 proteins overexpression. The reported results showed statistical significance ($p < 0.01$ **).

TDP-43 levels do not affect EPHX1 exon 5 processing

The EPHX1 (microsomal epoxide hydrolase 1) is a detoxifying enzyme that regulates oxidative stress. Interestingly, this enzyme can be linked to neuropathologies since significantly elevated levels of EPHX1 were found in hippocampal regions of patients suffering from Alzheimer's disease [342]. The bioinformatic analysis revealed a GU17 element in the downstream region of exon 5 (+20), which presents a very well defined consensus sequence ($CV=0.97=97\%$). Although EPHX1 levels were previously found to be downregulated upon TDP-43 silencing in a study from Baralle's laboratory [300], no alteration of the splicing pattern was observed in our minigene construct where exon 5 was always regularly included in all the different experimental condition tested (Figure 23a). Buratti et al. suggested that EPHX1 downregulation might be due to the specific upregulation of miR-663 expression upon TDP-43 removal [300]. Alternatively, TDP-43 might regulate EPHX1 at a transcriptional level or changes in EPHX1 levels might be ultimately caused by secondary modifications in the oxidative pathways occurring upon TDP-43 silencing.

TDP-43 levels do not affect NTRK1 exon 12 processing

As previously mentioned in Table 1, the NTRK1 gene encodes for the neurotrophic tyrosine kinase receptor type 1, also known as TRK1 or Trk-A (transforming tyrosine kinase protein 1/A). This kinase induces cell differentiation and might also play a role in specifying sensory neuron subtypes. Mutations in this gene have been associated with congenital insensitivity to pain with anhidrosis (CIPA), self-mutilating behavior, mental retardation and cancer [146, 147]. The bioinformatic analysis showed a GU16 element in the downstream region of exon 12 (+34). This UG repeat element has been reported as a polymorphic region, although the polymorphism was not found to correlate with incidence of CIPA disease (Table 1). The NTRK1 protein is almost ubiquitously expressed and different protein isoforms have been described. The NTRK1 exon 12 is reported as constitutively included in the final mRNA (Ensembl ENST00000534682) (CV=0=0%). In the NTRK1 exon 12 minigene system the exon 12 always resulted skipped and the splicing outcome was not affected by either TDP-43 silencing or overexpression (Figure 23b).

TDP-43 levels do not affect SGK2 exons 7, 8 processing

The SGK2 gene encodes for the serum-glucocorticoid regulated kinase 2, a modulator of the epithelial channels ENaC activity in human kidneys (ENaC gene has been previously reported in Table 1) [343, 344]. The SGK2 transcripts undergo alternative splicing leading to nine different protein isoforms (Ensembl ENSG00000101049). The bioinformatic analysis reported a long UGG(UG)18UA(UG)2 element at the 5'SS of SGK2 exon 8 (CV=0.67=67%). Given the particularly short exon 8 (37 nucleotides), in engineering the minigene construct to study the potential role of TDP-43 in splicing regulation I included also the upstream exon 7 (Figure 15A). In normal conditions, both exon 7 and 8 are generally included constitutively in SGK2 mRNA. In this study, SGK2 exon 8 was always included in the final mRNA and its processing was not affected by either TDP-43 silencing or overexpression (Figure 23c).

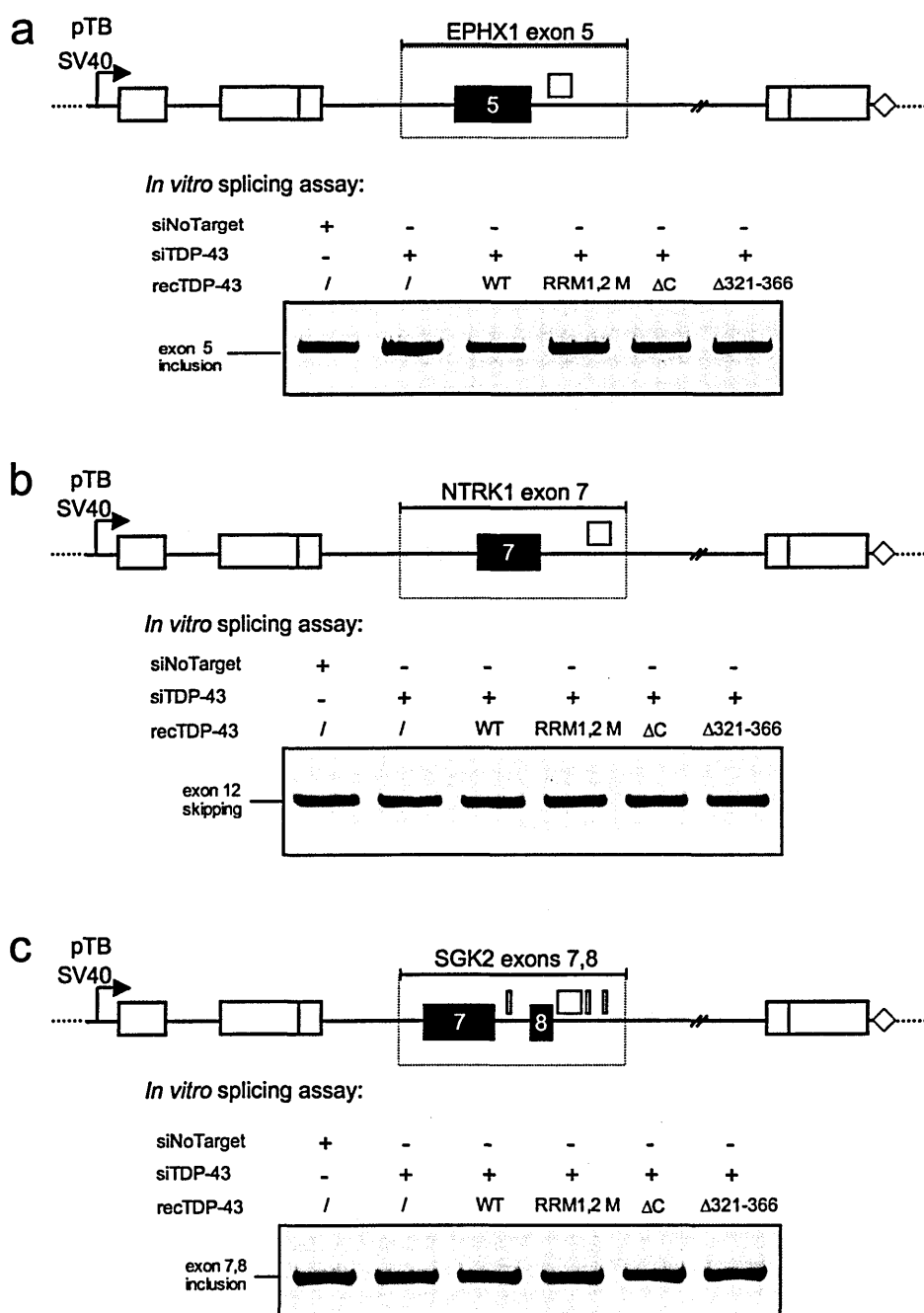


Figure 23. TDP-43 levels do not affect exon processing in EPHX1 exon 5, NTRK1 exon 12 and SGK2 exons 7, 8. a) A schematic representation of the EPHX1 exon 5 construct in the pTB reporter minigene used for splicing assay in HeLa cells and the splicing pattern unaffected by TDP-43 levels (N=2). b) A schematic representation of the NTRK1 exon 12 construct in the pTB reporter minigene used for splicing assay in HeLa cells and the splicing pattern unaffected by TDP-43 levels (N=2). c) A schematic representation of the SGK2 exons 7, 8 construct in the pTB reporter minigene used for splicing assay in HeLa cells and the splicing pattern unaffected by TDP-43 levels (N=2).

TDP-43 overexpression induces ETF1 exon 7 skipping

The ETF1 gene encodes for the eukaryotic translation termination factor 1, also called eukaryotic release factor 1 (eRF1). In eukaryotes ETF1 mediates translation termination in concert with eRF3/ETF3, as for the yeast homologues SUP45 and SUP35 [345]. In particular, ETF1 recognizes all the three termination codons (UAA, UGA, UAG) in the mRNAs inducing the hydrolysis of the peptidyl-tRNA ester bond and finally leading to translation termination [346, 347].

The bioinformatic analysis of the 70 nucleotides downstream region of an exon revealed the presence of a UG repeat element in the downstream sequence of ETF1 exon 7. However, additional analysis of the sequence revealed that the UG element extends beyond the 70 nucleotides reaching position +88. The UG rich element is composed by a UG₂ at +27 followed by an imperfect G(UG)₁₄UA(UG)₆UAUG repeated element at +41. The element is poorly conserved among primates and amniotes vertebrates (Figure 24b) and various ESE and ISE sequences compensate for the weak 5'SS consensus sequence (CV=0.23=23%) (Figure 30 in appendix).

Based on the complex pattern of alternative splicing involving ETF1 exon 7 (intron 7-8 retention and cryptic donor site within exon 7 upstream to the authentic one) (Ensembl ENSG00000120705; ETF1-008 ENST00000512198; ETF1-004 ENST00000506345) (Figure 24a), when engineering the ETF1 construct exon 8 was also included (Figure 25a).

Upon transfection in HeLa cells, in control conditions ETF1 exon 7 was mostly included and 33% skipped, whereas TDP-43 silencing enhanced exon 7 inclusion reporting only 5% of exon skipping (Figure 25b). This observation suggests that TDP-43 exerts an inhibitory effect on exon 7 processing. In accordance with this observation, overexpression of WT TDP-43 resulted in 73% of exon 7 skipping. Additionally, with RRM1,2M TDP-43 overexpression the

recover of exon 7 skipping did not occur, confirming once again that TDP-43 RNA-binding function is essential to mediate splicing modulation. Finally, skipping of ETF1 exon 7 was also observed when either ΔC or $\Delta 321-366$ TDP-43 were overexpressed, 60% and 47% respectively. The milder effect of $\Delta 321-366$ TDP-43 overexpression on exon 7 skipping in respect to ΔC TDP-43 suggests that the entire C-terminal regions might be required for TDP-43-mediated splicing regulation (Figure 25b).

Diagram illustrating alternative splicing of a pre-mRNA transcript. The top part shows a pre-mRNA with exons 5, 6, 7, 8, and 9, with introns 5-6, 6-7, 7-8, and 8-9. The bottom part shows three alternative splicing pathways:

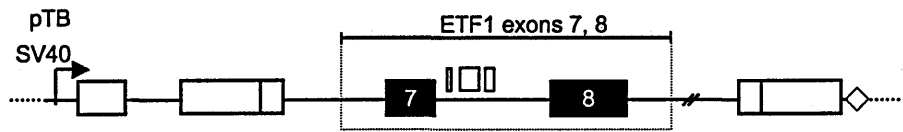
- 1) A mature mRNA with exons 5, 6, 7, 8, and 9.
- 2) An alternative 5'SS in exon 7, resulting in a mature mRNA with exons 5, 6, and 7.
- 3) Intron 7-8 retention + alternative 5'SS in exon 8, resulting in a mature mRNA with exons 5, 6, 7, and 8.

ETF 1

Human (<i>Homo sapiens</i>):	GTGAATGTTTTTCGTGTGTGTGTGTGTGTGTGTGTGTGTATGTGTGTGTGTATGTAC
Chimpanzee (<i>Pan troglodytes</i>):	GTGAATGTTTTTCGTGTGTGTGTGTGTGTGTGTGTGTGTG--TGTTGTGTGTGTGTGTAC
Gorilla (<i>Gorilla gorilla</i>):	GTGAATGTTTTTCGTGTGTGTGTA-----TATGTATGTATGTAC
Orangutan (<i>Pongo pygmaeus</i>):	GTGAATGTTTTTCGTGTGTGTGTGT-----GTATGTATGTATGTAC
Macaque (<i>Macaca mulatta</i>):	GTGAATGTTTTTCGTGTGTGTGTG-----GTACGTATATAC
Marmoset (<i>Callithrix jacchus</i>):	GTGAATGTTTTGTTTT--TG-----GAGGGGGTGTGTATGCATATGTATGTATGTAC
Chicken (<i>Gallus gallus</i>):	-----
Turkey (<i>Meleagris gallopavo</i>):	-----
Zebra Finch (<i>Taeniopygia guttata</i>):	-----
Mouse (<i>Mus musculus</i>):	-----TTTG-----
Rat (<i>Rattus norvegicus</i>):	-----TATG-----
Rabbit (<i>Oryctolagus cuniculus</i>):	-----TTTGTGTGTGT-----ATA--TGATGTAC
Horse (<i>Equus caballus</i>):	-----TTTATCTGTAT-----ATT--TGTGTGTAC
Dog (<i>Canis familiaris</i>):	-----TTTGTATGTAT-----A--TGCTGTAT
Pig (<i>Sus scrofa</i>):	-----TTTGTATGTGT-----
Cow (<i>Bos taurus</i>):	-----

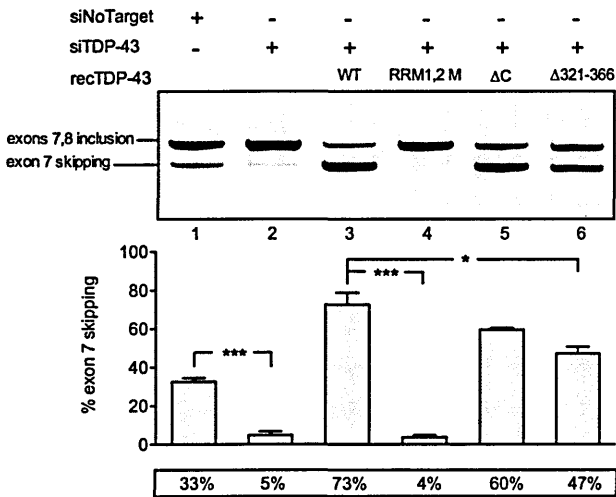
Figure 24. ETF1 exon 7 alternative processing and UG element conservation among mammals. a) ETF1 exon 7 can either be included or an alternative 5'SS within its sequence can be recognized. The resulting product in this case is reported to be degraded by NMD. Alternatively a third processed transcript with intron 7-8 retention and alternative 5'SS in exon 8 has been detected, but no corresponding protein product has been detected so far. b) The UG element at the 5'SS of exon 7 is poorly conserved among primates and amniotes vertebrates.

a ETF1 exons 7, 8 minigene



b

In vitro splicing assay:



c

Western blots:

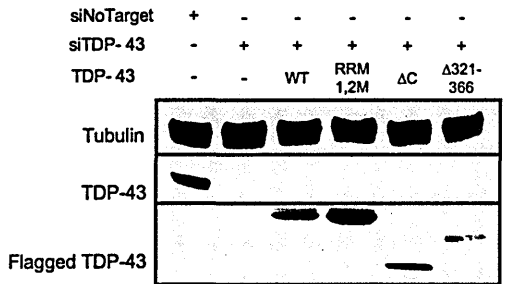


Figure 25. TDP-43 overexpression inhibits ETF1 exon 7 inclusion. a) A schematic representation of the ETF1 exon 7 construct in the pTB reporter minigene used for splicing assay in HeLa cells. b) Splicing assay in cell cultures reveals that TDP-43 silencing enhance exon 7 inclusion whereas TDP-43 overexpression inhibits it (N=3). c) Western blots detecting the TDP-43 silencing and flagged-tagged recombinant TDP-43 proteins overexpression. The asterisks represent the p values (p<0.05 *; p<0.01 **; p<0.001 ***) indicating the statistical significance of the results.

UG repeats enhance ETF1 exon 7 inclusion

Once proved that TDP-43 can affect the correct processing of ETF1 exon 7 through the binding of the UG repeats in the downstream region, the UG element was completely removed to observe the effect of UG repeats absence on exon 7 processing (Figure 26a). The Δ TG ETF1 construct was then transiently transfected in HeLa cells after TDP-43 silencing and overexpression.

The removal of the *cis*-regulatory element in the Δ TG ETF1 construct resulted in major exon skipping (70%), whereas in presence of the UG repeats exon 7 was mostly included (67%) (Figures 25b and 26b). These observations highlight the importance of the UG repeats as a splicing enhancer element ensuring the inclusion of the upstream exon 7. Upon either silencing or overexpression of TDP-43, exon 7 was still mostly skipped. In fact, in absence of the UG element TDP-43 cannot bind and therefore regulate the processing of the exon. As expected, overexpression of RRM1,2M, Δ C and Δ 321-366 TDP-43 proteins did not affect exon processing, given the absence of TDP-43 binding site (Figure 26b).

Interestingly, both saturation of the UG element with TDP-43 (in WT TDP-43 overexpression conditions, third lane on the gel) and removal of the UG element (in the Δ TG ETF1 construct in general) equally result in predominant exon 7 skipping (Figures 25b and 26b). These observations indicate that the long UG element itself represents an ESE promoting exon 7 inclusion likely through the recruitment of additional UG repeats-binding proteins, whereas TDP-43 act as an inhibitor of exon 7 inclusion. Consistent with this hypothesis, overexpression of TDP-43 possibly saturates the UG repeats element preventing the binding of other regulatory proteins with opposite (enhancing) effect ultimately resulting in exon 7 skipping. In line with this, TDP-43 silencing increased exon 7 inclusion (Figure 25b).

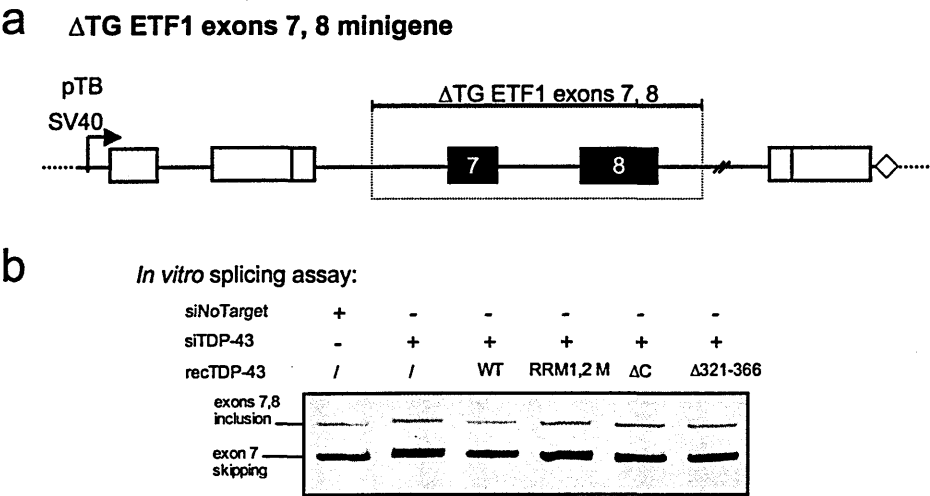


Figure 26. The UG repeated element enhances ETF1 exon 7 inclusion. a) A schematic representation of the ETF1 construct where the TG repeats have been deleted. b) Splicing assay in cell cultures reveals that TDP-43 does not affect exon 7 processing in absence of its UG repeats binding site (N=2).

In summary, TDP-43 does affect splicing of SGK2 exon 7-8, EPHX1 exon 5 and NTRK1 exon 12. However, in RXRG exon 7 and ETF1 exon 7 TDP-43 silences exon inclusion when overexpressed, therefore acting as a splicing inhibitor. In addition, the absence of UG repeats in the Δ TG ETF1 construct revealed that the UG element itself strongly promotes ETF1 exon 7 inclusion, further indicating the enhancer role of the UG repeats at the 5'SS in line with our previous results.

Potential TDP-43 involvement in endogenous 5'SS definition

In order to further establish whether the presence of 5'SS-located UG repeats can endogenously affect exon processing in a TDP-43-mediated manner, I started to analyze the endogenous splicing pattern of the exons reported in Table 3. Upon silencing of TDP-43 in HeLa cells, total RNA was extracted and semi-quantitative RT-PCR was performed in order to detect the splicing products. However, endogenous RNA analysis of 20 out of 45 exons did not show any splicing misregulation with this method (Figure 27a). Given the reported effect of TDP-43 overexpression in a SMN exon 7 minigene system [238], analysis of the endogenous SMN exon 7 splicing pattern was also performed, although no UG elements are present in the analyzed sequence. Also in this case no significant differences were observed upon TDP-43 depletion on SMN exon 7 endogenous processing.

Similarly, investigation of the endogenous effects of TDP-43 silencing on the five exons previously studied in the minigene systems did not report variations in the exons processing with this method (Figure 27b). Additionally, based on the results obtained with the ETF1 and Δ TG-ETF1 minigene constructs, the potential effect of both TDP-43 silencing and overexpression in the processing of endogenous ETF1 exon 7 were considered. To identify any possible splicing variation a wide genomic context was analyzed using a different set of oligonucleotides on ETF1 exons 5, 6, 7, 8 and 9 (Figure 28a). However, this approach did not reveal any variations in endogenous ETF1 exon 7 processing as well (Figure 28b).

a Endogenous RNA splicing assay

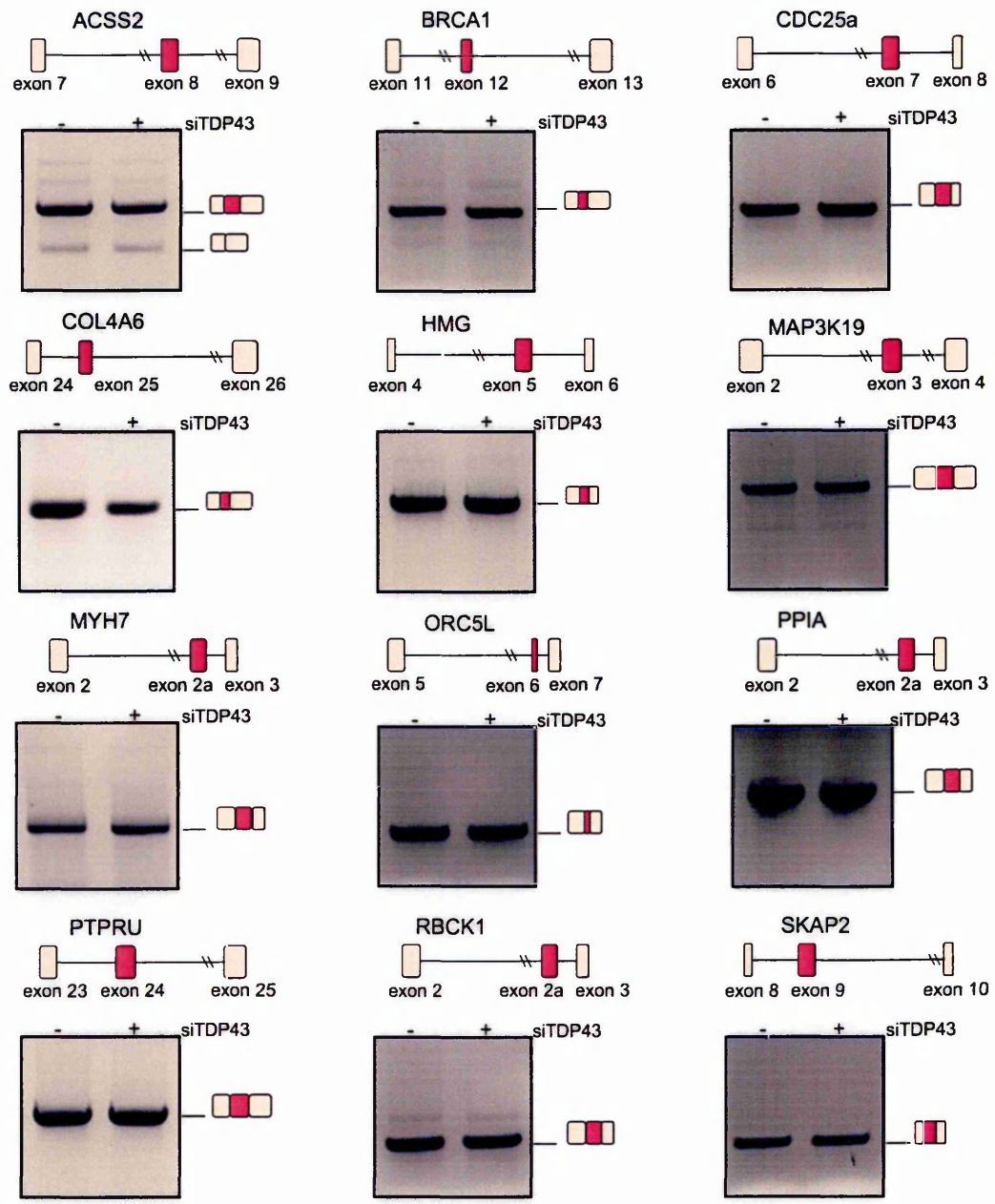


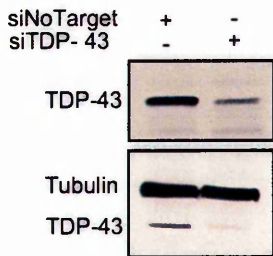
Figure 27a. TDP-43 silencing does not affect endogenous exon processing in the exons analyzed so far. For each studied exons a schematic representation of the analyzed sequence and the endogenous splicing assay are reported. The exons of interest are illustrated in purple color. Oligonucleotides on the flanking exons (pink color) have been used in PCR reactions to analyze the splicing patten of the central exon. For each exon the PCR reaction was performed twice using two endogenous RNA samples (N=2).

Endogenous RNA splicing assay:

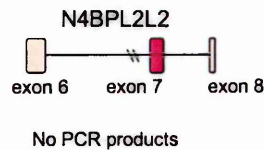
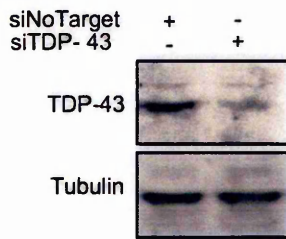


b Western Blots:

Trial 1



Trial 2



c Endogenous RNA splicing assay of the exons analyzed in minigene systems:

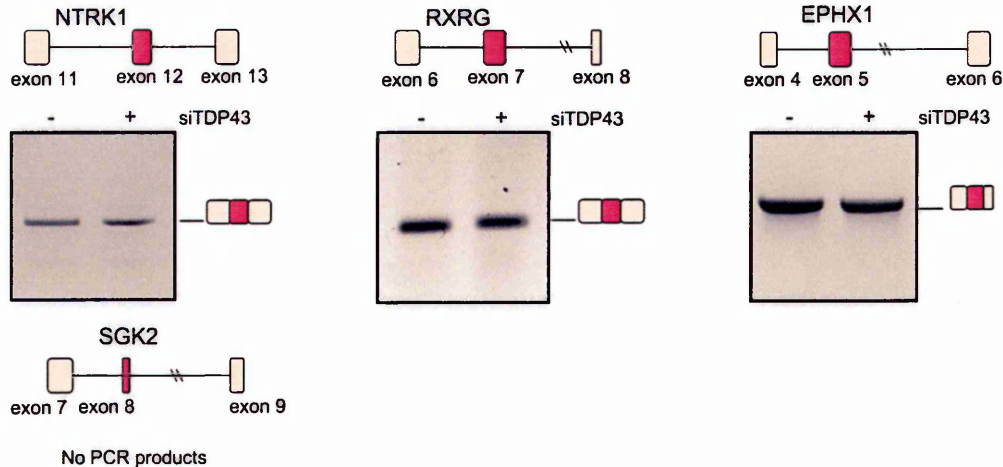


Figure 27a-c. TDP-43 silencing does not affect endogenous exon processing in the exons analyzed so far. For each studied exons a schematic representation of the analyzed sequence and the endogenous splicing assay are reported. For each exon the PCR reaction was performed twice using two endogenous RNA samples (Trial 1 and 2) (N=2). In some cases the PCR reaction did not report any product (although different oligonucleotides have been used) possibly because the corresponding mRNA is not expressed in HeLa. b) Western blots detecting the TDP-43 silencing in each trial are also reported. c) Endogenous analysis of the exons studied in minigene systems.

a Endogenous ETF1 RNA splicing assay:



b

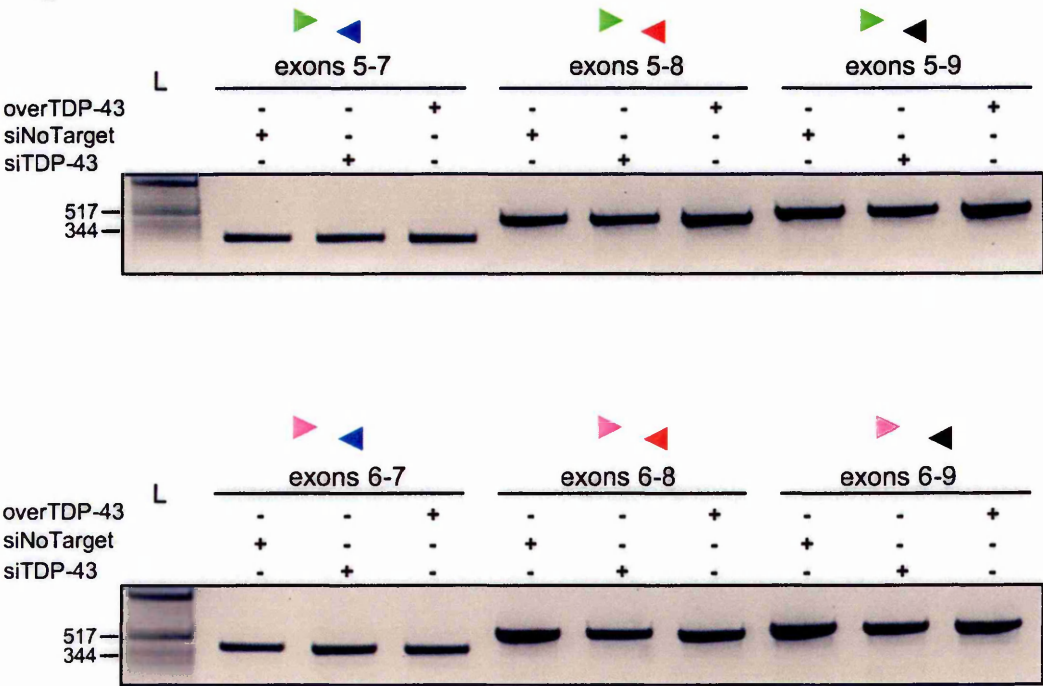


Figure 28. TDP-43 silencing and overexpression do not affect endogenous ETF1 exon 7 processing. a) A schematic representation of the wider genomic context of ETF1 considered. The small and different colored arrows represent the oligonucleotides used in PCR reactions. b) Endogenous RNA analysis did not show variations in ETF1 exon 7 processing in presence of TDP-43 silenced or overexpressed conditions (L: ladder) (N=2).

Discussion and Future Directions

Dinucleotide UG repeats located in proximity of splice site signals can act as additional *cis*-regulatory elements influencing exon processing in different splicing systems [88, 101, 161, 178, 279]. In particular, when located in proximity of the 5'SS of an exon the UG elements promote a particular event that consists in the exclusion of downstream exons, ultimately leading to shorter mRNA products [101, 161]. These UG repeats-mediated events result in tissue specific expression of the SCL8A1 gene [161] and promote an advantageous phenotype conferring resistance to malaria in the CD36 receptor [101]. This study further investigated the role of the 5'SS-located UG repeats by exploring their potential impact on splicing of the upstream exons. In particular, the synergic effect of the UG repeats-binding factor TDP-43 has been investigated. As shown in the results section, this has been mostly performed *in vitro* using several artificial and natural experimental systems using a variety of minigene-based techniques. Furthermore, analysis of the potential synergic effect of UG repeats and TDP-43 on the endogenous RNA processing has also begun to be explored.

***In vitro* minigene-based RNA analysis**

UG repeats and 5'SS definition

In the first part of this study, results obtained with an artificial system based on the pY7 reporter minigene suggested that the lone insertion of UG elements at the 5'SS does not significantly affect splicing of the upstream exon, regardless of different UG repeats length (6 and 12) and distances from the 5'SS (8 and 18 nucleotides) (Figure 8-10). In contrast to these results, it should be noted that this same reporter minigene was successfully used in the past to demonstrate that

longer UG repeats at the 3'SS inhibit splicing [277]. However, this initial experimental approach presented the limitation of analyzing the potential effect of UG repeats in a splicing-optimized system carrying a strong 5'SS consensus sequence.

Therefore, different natural systems carrying innate UG repeats in their 5'SS proximal intronic region were subsequently used as a more suitable substrate. In parallel, when evaluating the involvement of the UG repeats in splicing regulation, their ability to recruit UG repeats-binding proteins was also considered.

The first splicing systems to be analyzed was the BRCA1 exon 12 5'SS, which carries a GT6 at 9 nucleotides followed by a GT4 at 22 nucleotides from the 5'SS. This system presents the advantage of a natural and documented mutation that weakens the strength (in terms of consensus value) of the BRCA1 exon 12 5'SS, leading to exon 12 skipping *in vivo* in breast cancer patients [326].

The analysis of the mutated 5'SS of the BRCA1 exon 12 both in a two exons-one intron system (Figures 11-13) and the full length exon 12 in a three exons-two introns system (Figure 14) successfully mimicked the *in vivo* situation, confirming that the mutation is the direct cause of aberrant splicing events. Therefore, a competition assay was subsequently performed in order to investigate the involvement of UG repeats as *trans*-acting factors recruiter in BRCA1 exon 12 WT and MUT processing.

Splicing analysis of the BRCA1 exon 12 5'SS WT in the presence of GU6 RNA competing oligos revealed that the UG elements at the 5'SS of exon 12 can enhance exon processing through the recruitment of UG repeats-binding proteins (Figure 15a). Interestingly, this enhancement was strongly marked in the BRCA1 exon 12 5'SS MUT carrying the weakened 5'SS (Figure 15b). Generally, additional regulatory elements are required for proper exon processing in the presence of weak canonical splicing signals (5'SS and 3'SS) [348, 349]. Consistent with this

knowledge, our results indicated that UG repeats and UG repeats-binding proteins are indeed required for proper 5'SS definition of the poorly defined exon 12. Further analysis showed that in this system the UG repeats-binding proteins also favour the transition between the A complex and late complexes, possibly by recruiting additional spliceosomal components, as shown by the analysis of the spliceosomal complex assembly (Figure 16).

It can be noted that the 5'SS strength-dependent synergic role played by the UG repeats and UG repeats-binding proteins in 5'SS definition observed in the BRCA1 exon 12 5'SS MUT system resembles the genomic context-dependent effect observed in the APOAII splicing system by Mercado et al.. In fact, polymorphic UG repeats at the 3'SS were found to induce APOAII exon 3 skipping only upon disruption of additional regulatory sequences with inclusion-enhancing properties [88]. Similarly, in the BRCA1 exon 12 model the UG repeats at the 5'SS play a key role in 5'SS definition, through the recruitment of UG repeats-binding proteins, but only when the strong WT 5'SS is weakened by the mutation.

In addition to these considerations on the effects of UG repeats on 5'SS recognition it should be noted that, from a simple diagnostic point of view, these conclusions represent the first mechanistic evaluation of a putative splicing-affecting pathological mutation, providing insight of a direct correlation between the genetic variation and the splicing defect.

TDP-43 enhances mutated BRCA1 exon 12 inclusion

Among the numerous UG repeats-binding proteins reported in literature (PTB, FUBP3, CELF1, CELF2, SRp55, etc.), we have focused predominantly on TDP-43, a protein with great affinity for this sequence [198]. TDP-43 is ubiquitously expressed in human tissues and has been described to be involved in neuropathologies such as ALS and FTLT. In this study we investigated TDP-43 involvement in the regulation of different splicing systems.

TDP-43 is bona fide a heterogeneous nuclear ribonucleoprotein acting as a splicing regulator (either inhibitor or activator) [86, 88, 178, 238, 239, 241], as well as influencing other post-transcriptional steps in gene expression [199, 240, 242, 243, 268, 270, 300]. Since our studies on the BRCA1 exon 12 system indicated that UG repeats-binding proteins in general help the processing of the exon, the role played by TDP-43 in this splicing system was investigated. Experiments in HeLa cells showed that either TDP-43 silencing or overexpression failed to affect the splicing of the well defined BRCA1 exon 12 5'SS WT. However, TDP-43 silencing completely inhibited proper exon 12 recognition in presence of the weak mutated 5'SS in BRCA1 exon 12 MUT, whereas overexpression of the protein partially restored exon 12 inclusion (Figure 18). These results indicated that the UG repeats at the 5'SS of BRCA1 exon 12 MUT can lead to TDP-43-mediated enhancement of exon 12 inclusion. Hence, our investigation allowed the identification of novel system in which the protein displays a 5'SS strength-dependent splicing enhancing functionality.

Interestingly, quite a few other cases of TDP-43-mediated splicing activation are found in the literature. The first case was reported by Bose et al. in 2008 and concerned the SMN2 exon 7 system, where a mutation in position +6 of exon 7 leads to major exon skipping. In this study, whereas silencing of TDP-43 did not affect exon 7 processing at all, overexpression of TDP-43 successfully restored

exon 7 inclusion, although no UG elements were present in this system [238]. Most recently, Tollervey et al. reported that TDP-43 silencing enhances alternative skipping of CDK5RAP2 exon 39 and GPBP1L1 exon 4, which carry UG repeats in their “deep” intronic sequences, implying TDP-43 correlation with alternative exon inclusion enhancement in normal condition [241]. Similarly, seven exons were found mostly skipped following TDP-43 depletion in mouse brain: POLDIP3 exon 3, TSN exon 5, RBM33 exon 6, DCLK1 exon 16, DST exon 96 and PPP3ACA exon 13. However, since in these cases no significant TDP-43 binding sites were found within 2kb of the splice sites of most of the altered splicing targets the authors suggested an indirect effect of TDP-43, possibly through the alteration of other splicing factors [86].

Considerations on TDP-43 binding sites and splicing modulation activity

In the second part of this study, a genome-wide bioinformatic analysis brought to the identification of 45 exons carrying long UG repeats elements in their 5'SSs or proximal intronic sequences (within 70 nucleotides from the 5'SS). Given the presence of UG repeats, these exons may represent potential targets of the UG repeats-binding protein TDP-43. Consistent with this simple assumption is the presence of UG motifs in most of the recently identified TDP-43 RNA targets [236, 241].

Among the 45 identified exons, we selected for further studies five exons with reported TDP-43 clusters on CLIP analysis (Prof. Jernej Ule, personal communication) and whose transcripts levels were affected by TDP-43 silencing in our microarray analysis [199]. Splicing assay analysis in HeLa cells revealed that three out of five examined exons (NTRK1 exon 12, EPHX1 exon 5 and SGK2 exon 7, 8) were not affected by either silencing or overexpression of TDP-43. These observations suggested that either other RNA-binding proteins successfully

compensate TDP-43 altered levels or, alternatively, that these exons are not bound by TDP-43 although the presence of its preferential binding site (Figure 23). Consistent with the latter, recently published CLIP data claimed that 96% of RNAs containing UG-clusters are actually unaffected by TDP-43 depletion *in vivo* in mouse [86], indicating that not all the UG motifs necessarily represent a TDP-43 binding site.

On the other hand, in ETF1 exon 7 and RXRG exon 7 systems, overexpression of TDP-43 resulted in skipping of the corresponding exons, identifying a previously unknown dependency on TDP-43 of these two splicing events *in vitro* (Figures 22 and 25). Additionally, in the ETF1 exon 7, silencing of TDP-43 produced the expected opposite effect of enhancing exon 7 inclusion (Figure 26). The inhibitory role played by TDP-43 was first described in early studies in Baralle's laboratory showing that UG repeats at the 3'SS of the CFTR exon 9 and APOAII exon 3 inhibit exon inclusion [88, 173]. More recently, UG repeats in the upstream proximal intronic sequence of ZNF92 exon 2 and PILRB exon 39 and in the downstream proximal intronic sequence of REEP6 exon 5 and CACNA1C exon 13 were also found to promote skipping of the corresponding proximal exons through the recruitment of TDP-43 [241]. In addition, nine exons were found to be mostly included upon TDP-43 silencing in mouse (SORT1 exon 18, KCND3 exon 5, MPZL1 exon 5, DDEF2 exon 25, PDP1 exon 1, SEMA3F exon 5, ATP2B2 exons 6, 7, 8, EIF4H exon 5, DNAJC5 exon 4a) [86].

In conclusion, in line with the previously described cases of TDP-43-mediated silencing of exon inclusion, the results obtained in the ETF1 exon 7 and RXRG exon 7 reported two additional systems depending on TDP-43 levels.

TDP-43 overexpression silences ETF1 exon 7 inclusion

The downstream sequence of the human ETF1 exon 7 present a long, poorly conserved throughout evolution and nearly perfect UG element (Figure 24). By investigating the role of TDP-43 in exon 7 regulation through the binding of these TG repeats we found that silencing of TDP-43 enhances ETF1 exon 7 inclusion, whereas its overexpression results in exon 7 skipping (Figure 25). In addition, the removal of the UG repeats in the Δ TG ETF1 construct mostly resulted in exon 7 skipping, implying that the UG repeats *per se* enhance exon inclusion, possibly by recruiting additional splicing factors (Figure 26). In this case, TDP-43 might exert its inhibitory activity by competing for the binding site with other UG repeats-binding proteins with splicing-enhancing properties. Moreover, the fact that either TDP-43 silencing or overexpression did not affect exon 7 splicing in the absence of UG repeats (Δ TG ETF1 construct) supports the direct involvement of TDP-43 in splicing regulation through these identified UG repeats, rather than through an indirect effect caused by its eventual modulation of other splicing factors (Figure 26).

In fact, the potential indirect effect of this protein on general splicing regulation is supported by the fact that TDP-43 binding sites were identified on pre-mRNAs of different RNA-binding proteins directly involved in the splicing process (FUS/TLS, TAF15, Nova1 and 2, Mbnl, nPTB, etc.). In these instances, either the mRNA levels of these proteins were altered or alternatively spliced upon TDP-43 silencing (i.e. FUS/TLS mRNA reduced levels; nPTB exon 10 skipping; TIA1 exon 5 inclusion) [86]. Based on these observations, and in light of the enormous amount of genes (over 6000) found affected by TDP-43 levels, a likely indirect effect of TDP-43 has to be considered when analyzing a splicing systems affected by either TDP-43 silencing or overexpression.

In conclusion, from a functional point of view it can be reasoned that by affecting the ETF1 mRNA transcripts, a key factor in translation termination, TDP-43 might indirectly affect translation of several proteins, regulating their levels post-transcriptionally. The next stage of this work will be to validate this hypothesis *in vivo* in order to obtain a clear evaluation of the real functional consequences of exon 7 regulation by TDP-43.

TDP-43 overexpression silences RXRG exon 7 inclusion

The RXRG exon 7 downstream sequence presents a long imperfect UG element, mostly conserved among primates, but poorly conserved among amniotes vertebrates (Figure 21). Upon overexpression of TDP-43, exon 7 is nearly 30% skipped (Figure 22). Since in normal minigene conditions exon 7 is completely included in the final mRNA, no enhancing effect on exon inclusion expected upon TDP-43 silencing could be detected with this approach. As reported for the ETF1 exon 7, TDP-43 might generally interfere with the enhancing effect of the UG repeats at the 5'SS, ultimately silencing RXRG exon 7 inclusion. Also in this case, both direct and indirect effects of TDP-43 will have to be considered in further studies in order to better define TDP-43 involvement in this splicing event and its consequences.

The recent discovery of its key function in accelerating spontaneous remyelination in multiple sclerosis lesions has brought great scientific interest to the RXRG gene, which might represent a pharmacological target for regenerative therapy in the central nervous system [350]. Interestingly, Latasa et al. recently reported that retinoic acid (RA) strongly inhibits RXRG expression [351]. Interestingly, and as already presented in the introduction, the majority of the RA-downregulated genes present TG repeats in their regulatory sequences. These TG repeats are required for RA-induced repression of gene expression in

Xenopus laevis embryos through the recruitment of PTB, another TG repeats-binding protein [171, 172]. Based on these published data and on our results on RXRG exon 7, it would therefore be interesting to investigate whether TDP-43-mediated exon 7 skipping might interfere with the capability of this RA receptor to bind its ligand and to trigger RA-induced gene modulation. In fact, RXRG exon 7, together with exons 6, 8, 9 and 10 is expected to code for the protein ligand-binding domain (information not based on experimental findings, but transferred by similarity to other protein family members). Furthermore, TDP-43 indirect role on RXRG mRNA processing through its direct modulation of the splicing factor nPTB (neuronal PTB) might be a possible explanation that needs further investigation [86].

Also in this case, further work will be aimed to test the effects of TDP-43 on RXRG exon 7 at the endogenous level in different experimental conditions to accurately determine its functional consequences *in vivo*.

Better understanding of TDP-43 functionality

Interestingly, a growing number of RNA-binding proteins were found to exert a position-dependent regulation of splicing events, such as Fox-1 [352], PTB [353], Nova, Mbnl1, Sxl, hnRNP C, L and H [354]. Generally, when the corresponding binding sites are located in the upstream proximal intronic regions or in the 5' and 3'SSs, the RNA-binding proteins silence exon inclusion; alternatively they enhance exon inclusion when binding at the downstream region of an exon [354]. In line with these observations, a recent study from Tollervey et al. proposed a position-dependent regulation of splicing for TDP-43 as well. Although several exceptions were reported in this same study, the authors identified a tendency for TDP-43 to silence exon inclusion when binding UG repeats located within the exons, -1 to -150 nucleotides upstream or +150 to +500

nucleotides downstream (deep intronic sequences), but to enhance exon inclusion when binding +1 to +150 nucleotides downstream of alternative exons [241].

Our *in vitro* results, however, contrast with these conclusions because TDP-43 binding to UG repeats within +70 nucleotides resulted in silencing of exon inclusion both in ETF1 exon 7 and RXRG exon 7 systems (UG repeats at +36 and +42 respectively). Alternatively, our findings highlight a possible correlation between the effects on splicing of TDP-43 and the splice site features, in particular the 5'SS strength, rather than a correlation with the position of the binding site. This observation is supported by the fact that both TDP-43-repressed exons (ETF1 exon 7 and RXRG exon 7) and TDP-43-activated exon (BRCA1 MUT) present weak 5'SS, whereas unaffected exons present strong 5'SS values (BRCA1 WT, SKG2, EPHX1). Therefore, based on these data it can be reasoned that TDP-43 levels are an important factor defining weak 5'SSs.

Finally, our analysis of the different TDP-43 domains involvement in splicing regulation indicated that in all cases (BRCA1 MUT, RXRG and ETF1) the integrity of the RNA-binding domains is absolutely required for TDP-43 to affect splicing (Figure 12, 14, 18), further confirming pre-existing data [203]. Mutation of these domains results in the abolishment of TDP-43 splicing functionality. In addition, overexpression of the proteins lacking the entire C-terminal domain or the 321-366 domain always mimicked the overexpression of the WT protein, although with weaker penetrance, indicating that the C-terminal domain of the protein is important but not strictly necessary for the protein to exert its splicing regulatory functions.

Endogenous RNA analysis

Following the results of these *in vitro* studies the logical follow up of this work has been to evaluate the effects of TDP-43 depletion in the endogenous context. Disappointingly, a preliminary analysis of the endogenous RNAs in TDP-43 silenced conditions did not reveal any splicing variation for all the tested genes with putative UG-regulatory motifs near their 5'SSs (Figure 27). Of course this was particularly disappointing for the RXRG and ETF1 genes, where the analysis of the endogenous RNA did not report the expected results, although a neat effect of TDP-43 levels on exon processing was observed in minigene analysis.

However, there are several possibilities that could account for this apparent discrepancy. In fact, the splicing pattern observed in the two minigene systems could be representative of a tissue-specific event, yet to be characterized. For example, RXRG is highly expressed in brain and skeletal muscle, whereas we analyzed its endogenous splicing pattern in cervix-derived cells (HeLa). Therefore, in order to test this possibility it will be important to evaluate in the future the endogenous RXRG expression in these tissues following TDP-43 depletion. In addition, the developmental stage could be an important factor determining the effects of TDP-43 in the inclusion of these exons, and this possibility can be investigated using suitable cellular and animal models. Finally, the combination of quality and quantity of splicing factors present in natural versus transfected cell environment could also negatively influence the role of TDP-43 in a particular endogenous situation making it less determinant.

Nevertheless, it is important to note that a comparison between minigenes and endogenous situations always requires some careful considerations. Not necessarily, in fact, an apparent lack of consistency between these two systems should be taken as a demonstration that minigene results do not reflect the endogenous situation. Regarding this issue, Baralle's laboratory has clearly shown

in the past that apparent discrepancies can sometimes be solved by a careful consideration of the various factors that are known to affect the splicing process. For example, it has been shown that in the NF1 (neurofibromatosis 1) gene the genomic context plays a very important role in explaining apparent discrepancies between minigenes and endogenous situations [355]. In particular, it was observed that the minigene system did not successfully mimic the endogenous splicing pattern unless an extended genomic context was added. Because our situation is different (an effect on splicing was observed only in the minigene context but not at the endogenous level), it was impossible to try and solve these discrepancies by performing a similar analysis on the ETF1 and RXGR systems. Nonetheless, what has been seen with the minigene systems is a potential effect of TDP-43 on exon recognition of these two genes. Whether this potential influence may be confined to particular cell types/developmental stages/combination of splicing factors, as discussed above, is something that will require further studies (that are currently in progress).

In conclusion, this study has provided positive proof of concept that UG repeats located in the downstream region of poorly defined exons may help 5'SS definition, acting as auxiliary *cis*-regulatory elements *in vitro*. In addition, this study has improved our understanding of TDP-43 involvement in splicing regulation with the identification and characterization of three potential and previously unknown RNA targets. Also, in line with the most recent findings [86, 241], this study has shown that, *in vitro*, TDP-43 can either function as a splicing inhibitor or activator.

References

1. Proudfoot, N., *New perspectives on connecting messenger RNA 3' end formation to transcription*. Curr Opin Cell Biol, 2004. **16**(3): p. 272-8.
2. Hirose, Y. and J.L. Manley, *RNA polymerase II and the integration of nuclear events*. Genes Dev, 2000. **14**(12): p. 1415-29.
3. Orphanides, G. and D. Reinberg, *A unified theory of gene expression*. Cell, 2002. **108**(4): p. 439-51.
4. Reed, R., *Coupling transcription, splicing and mRNA export*. Curr Opin Cell Biol, 2003. **15**(3): p. 326-31.
5. Kornblihtt, A.R., et al., *Multiple links between transcription and splicing*. RNA, 2004. **10**(10): p. 1489-98.
6. Lund, M. and J. Kjems, *Defining a 5' splice site by functional selection in the presence and absence of U1 snRNA 5' end*. RNA, 2002. **8**(2): p. 166-79.
7. Roca, X., R. Sachidanandam, and A.R. Krainer, *Intrinsic differences between authentic and cryptic 5' splice sites*. Nucleic Acids Res, 2003. **31**(21): p. 6321-33.
8. Reed, R. and T. Maniatis, *Intron sequences involved in lariat formation during pre-mRNA splicing*. Cell, 1985. **41**(1): p. 95-105.
9. Reed, R. and T. Maniatis, *The role of the mammalian branchpoint sequence in pre-mRNA splicing*. Genes Dev, 1988. **2**(10): p. 1268-76.
10. Reed, R., *The organization of 3' splice-site sequences in mammalian introns*. Genes Dev, 1989. **3**(12B): p. 2113-23.
11. Matlin, A.J., F. Clark, and C.W. Smith, *Understanding alternative splicing: towards a cellular code*. Nat Rev Mol Cell Biol, 2005. **6**(5): p. 386-98.
12. Gregory, T.R. and P.D. Hebert, *The modulation of DNA content: proximate causes and ultimate consequences*. Genome Res, 1999. **9**(4): p. 317-24.
13. Lander, E.S., et al., *Initial sequencing and analysis of the human genome*. Nature, 2001. **409**(6822): p. 860-921.
14. Ohno, S., *So much "junk" DNA in our genome*. Brookhaven Symp Biol, 1972. **23**: p. 366-70.
15. Orgel, L.E. and F.H. Crick, *Selfish DNA: the ultimate parasite*. Nature, 1980. **284**(5757): p. 604-7.
16. Richard, G.F., A. Kerrest, and B. Dujon, *Comparative genomics and molecular dynamics of DNA repeats in eukaryotes*. Microbiol Mol Biol Rev, 2008. **72**(4): p. 686-727.
17. Gemayel, R., et al., *Variable tandem repeats accelerate evolution of coding and regulatory sequences*. Annu Rev Genet, 2010. **44**: p. 445-77.
18. Ellegren, H., *Microsatellites: simple sequences with complex evolution*. Nat Rev Genet, 2004. **5**(6): p. 435-45.
19. Hui, J., G. Reither, and A. Bindereif, *Novel functional role of CA repeats and hnRNP L in RNA stability*. RNA, 2003. **9**(8): p. 931-6.
20. Sharma, V.K., et al., *Abundance of dinucleotide repeats and gene expression are inversely correlated: a role for gene function in addition to intron length*. Physiol Genomics, 2007. **31**(1): p. 96-103.
21. Huichalaf, C., et al., *Expansion of CUG RNA repeats causes stress and inhibition of translation in myotonic dystrophy 1 (DM1) cells*. FASEB J, 2010. **24**(10): p. 3706-19.
22. Gau, B.H., et al., *FUBP3 interacts with FGF9 3' microsatellite and positively regulates FGF9 translation*. Nucleic Acids Res, 2011.
23. Riley, D.E. and J.N. Krieger, *Short tandem repeats are associated with diverse mRNAs encoding membrane-targeted proteins*. Bioessays, 2004. **26**(4): p. 434-44.
24. Johnson, A.C., et al., *Epidermal growth factor receptor gene promoter. Deletion analysis and identification of nuclear protein binding sites*. J Biol Chem, 1988. **263**(12): p. 5693-9.

25. Contente, A., et al., *A polymorphic microsatellite that mediates induction of PIG3 by p53*. Nat Genet, 2002. **30**(3): p. 315-20.
26. Johnson, A.C., Y. Jinno, and G.T. Merlino, *Modulation of epidermal growth factor receptor proto-oncogene transcription by a promoter site sensitive to S1 nuclease*. Mol Cell Biol, 1988. **8**(10): p. 4174-84.
27. van Ham, S.M., et al., *Phase variation of H. influenzae fimbriae: transcriptional control of two divergent genes through a variable combined promoter region*. Cell, 1993. **73**(6): p. 1187-96.
28. Frisch, R., et al., *Effect of triplet repeat expansion on chromatin structure and expression of DMPK and neighboring genes, SIX5 and DMWD, in myotonic dystrophy*. Mol Genet Metab, 2001. **74**(1-2): p. 281-91.
29. Rothenburg, S., et al., *A polymorphic dinucleotide repeat in the rat nucleolin gene forms Z-DNA and inhibits promoter activity*. Proc Natl Acad Sci U S A, 2001. **98**(16): p. 8985-90.
30. Potaman, V.N., et al., *Length-dependent structure formation in Friedreich ataxia (GAA)_n*(TTC)_n repeats at neutral pH*. Nucleic Acids Res, 2004. **32**(3): p. 1224-31.
31. Hefferon, T.W., et al., *A variable dinucleotide repeat in the CFTR gene contributes to phenotype diversity by forming RNA secondary structures that alter splicing*. Proc Natl Acad Sci U S A, 2004. **101**(10): p. 3504-9.
32. Tian, B., et al., *Expanded CUG repeat RNAs form hairpins that activate the double-stranded RNA-dependent protein kinase PKR*. RNA, 2000. **6**(1): p. 79-87.
33. Stern, A., et al., *Opacity genes in Neisseria gonorrhoeae: control of phase and antigenic variation*. Cell, 1986. **47**(1): p. 61-71.
34. Weiser, J.N., J.M. Love, and E.R. Moxon, *The molecular mechanism of phase variation of H. influenzae lipopolysaccharide*. Cell, 1989. **59**(4): p. 657-65.
35. Hoyer, L.L., *The ALS gene family of Candida albicans*. Trends Microbiol, 2001. **9**(4): p. 176-80.
36. Verstrepen, K.J., et al., *Intragenic tandem repeats generate functional variability*. Nat Genet, 2005. **37**(9): p. 986-90.
37. Sawyer, L.A., et al., *The period gene Thr-Gly polymorphism in Australian and African Drosophila melanogaster populations: implications for selection*. Genetics, 2006. **174**(1): p. 465-80.
38. Lee, K., J.C. Dunlap, and J.J. Loros, *Roles for WHITE COLLAR-1 in circadian and general photoperception in Neurospora crassa*. Genetics, 2003. **163**(1): p. 103-14.
39. Sawyer, L.A., et al., *Natural variation in a Drosophila clock gene and temperature compensation*. Science, 1997. **278**(5346): p. 2117-20.
40. Johnsen, A., et al., *Avian Clock gene polymorphism: evidence for a latitudinal cline in allele frequencies*. Mol Ecol, 2007. **16**(22): p. 4867-80.
41. Legendre, M., et al., *Sequence-based estimation of minisatellite and microsatellite repeat variability*. Genome Res, 2007. **17**(12): p. 1787-96.
42. Fondon, J.W., 3rd and H.R. Garner, *Molecular origins of rapid and continuous morphological evolution*. Proc Natl Acad Sci U S A, 2004. **101**(52): p. 18058-63.
43. La Spada, A.R., et al., *Androgen receptor gene mutations in X-linked spinal and bulbar muscular atrophy*. Nature, 1991. **352**(6330): p. 77-9.
44. Orr, H.T. and H.Y. Zoghbi, *Trinucleotide repeat disorders*. Annu Rev Neurosci, 2007. **30**: p. 575-621.
45. Kuyumcu-Martinez, N.M., G.S. Wang, and T.A. Cooper, *Increased steady-state levels of CUGBP1 in myotonic dystrophy 1 are due to PKC-mediated hyperphosphorylation*. Mol Cell, 2007. **28**(1): p. 68-78.
46. Miller, J.W., et al., *Recruitment of human muscleblind proteins to (CUG)_n expansions associated with myotonic dystrophy*. EMBO J, 2000. **19**(17): p. 4439-48.
47. Khani-Hanjani, A., et al., *Identification of four novel dinucleotide repeat polymorphisms in the TNF-alpha and TNF-beta genes*. Hum Immunol, 2000. **61**(5): p. 511-2.
48. Hiraoka, M., et al., *Insertion and deletion mutations in the dinucleotide repeat region of the Norrie disease gene in patients with advanced retinopathy of prematurity*. J Hum Genet, 2001. **46**(4): p. 178-81.

49. Hui, J., et al., *Intronic CA-repeat and CA-rich elements: a new class of regulators of mammalian alternative splicing*. EMBO J, 2005. **24**(11): p. 1988-98.
50. Xiao, X., et al., *Splice site strength-dependent activity and genetic buffering by poly-G runs*. Nat Struct Mol Biol, 2009. **16**(10): p. 1094-100.
51. McCullough, A.J. and S.M. Berget, *An intronic splicing enhancer binds U1 snRNPs to enhance splicing and select 5' splice sites*. Mol Cell Biol, 2000. **20**(24): p. 9225-35.
52. Marcucci, R., F.E. Baralle, and M. Romano, *Complex splicing control of the human Thrombopoietin gene by intronic G runs*. Nucleic Acids Res, 2007. **35**(1): p. 132-42.
53. Russo, A., et al., *hnRNP H1 and intronic G runs in the splicing control of the human rpl3 gene*. Biochim Biophys Acta, 2010. **1799**(5-6): p. 419-28.
54. Lu, Z.X., et al., *Context-dependent robustness to 5' splice site polymorphisms in human populations*. Hum Mol Genet, 2011. **20**(6): p. 1084-96.
55. Hui, J., et al., *HnRNP L stimulates splicing of the eNOS gene by binding to variable-length CA repeats*. Nat Struct Biol, 2003. **10**(1): p. 33-7.
56. Rothrock, C.R., A.E. House, and K.W. Lynch, *HnRNP L represses exon splicing via a regulated exonic splicing silencer*. EMBO J, 2005. **24**(15): p. 2792-802.
57. Heiner, M., et al., *HnRNP L-mediated regulation of mammalian alternative splicing by interference with splice site recognition*. RNA Biol, 2010. **7**(1): p. 56-64.
58. Lee, D.H., et al., *hnRNP L binds to CA repeats in the 3'UTR of bcl-2 mRNA*. Biochem Biophys Res Commun, 2009. **382**(3): p. 583-7.
59. Rietveld, I., et al., *A polymorphism in the IGF-I gene influences the age-related decline in circulating total IGF-I levels*. Eur J Endocrinol, 2003. **148**(2): p. 171-5.
60. Wu, T., et al., *Characterization of the promoter for the human 85 kDa cytosolic phospholipase A2 gene*. Nucleic Acids Res, 1994. **22**(23): p. 5093-8.
61. Borrmann, L., et al., *Human HMGA2 promoter is coregulated by a polymorphic dinucleotide (TC)-repeat*. Oncogene, 2003. **22**(5): p. 756-60.
62. Wang, B., et al., *Dinucleotide repeats negatively modulate the promoter activity of Cyr61 and is unstable in hepatocellular carcinoma patients*. Oncogene, 2005. **24**(24): p. 3999-4008.
63. Gebhardt, F., K.S. Zanker, and B. Brandt, *Modulation of epidermal growth factor receptor gene transcription by a polymorphic dinucleotide repeat in intron 1*. J Biol Chem, 1999. **274**(19): p. 13176-80.
64. Philips, A.V., L.T. Timchenko, and T.A. Cooper, *Disruption of splicing regulated by a CUG-binding protein in myotonic dystrophy*. Science, 1998. **280**(5364): p. 737-41.
65. Mankodi, A., et al., *Expanded CUG repeats trigger aberrant splicing of CIC-1 chloride channel pre-mRNA and hyperexcitability of skeletal muscle in myotonic dystrophy*. Mol Cell, 2002. **10**(1): p. 35-44.
66. Jensen, K.B., et al., *The tetranucleotide UCAY directs the specific recognition of RNA by the Nova K-homology 3 domain*. Proc Natl Acad Sci U S A, 2000. **97**(11): p. 5740-5.
67. Ule, J., et al., *CLIP identifies Nova-regulated RNA networks in the brain*. Science, 2003. **302**(5648): p. 1212-5.
68. Dredge, B.K., et al., *Nova autoregulation reveals dual functions in neuronal splicing*. EMBO J, 2005. **24**(8): p. 1608-20.
69. Hedjran, F., et al., *Control of alternative pre-mRNA splicing by distributed pentameric repeats*. Proc Natl Acad Sci U S A, 1997. **94**(23): p. 12343-7.
70. Huh, G.S. and R.O. Hynes, *Regulation of alternative pre-mRNA splicing by a novel repeated hexanucleotide element*. Genes Dev, 1994. **8**(13): p. 1561-74.
71. Lynch, K.W. and T. Maniatis, *Assembly of specific SR protein complexes on distinct regulatory elements of the Drosophila doublesex splicing enhancer*. Genes Dev, 1996. **10**(16): p. 2089-101.
72. Sharma, V.K., et al., *(TG:CA)(n) repeats in human housekeeping genes*. J Biomol Struct Dyn, 2003. **21**(2): p. 303-10.

73. Sharma, V.K., S.K. Brahmachari, and S. Ramachandran, *(TG/CA)_n repeats in human gene families: abundance and selective patterns of distribution according to function and gene length*. BMC Genomics, 2005. 6: p. 83.
74. Bayele, H.K., et al., *HIF-1 regulates heritable variation and allele expression phenotypes of the macrophage immune response gene SLC11A1 from a Z-DNA forming microsatellite*. Blood, 2007. 110(8): p. 3039-48.
75. Haniford, D.B. and D.E. Pulleyblank, *Facile transition of poly[d(TG) x d(CA)] into a left-handed helix in physiological conditions*. Nature, 1983. 302(5909): p. 632-4.
76. Nordheim, A. and A. Rich, *The sequence (dC-dA)_n X (dG-dT)_n forms left-handed Z-DNA in negatively supercoiled plasmids*. Proc Natl Acad Sci U S A, 1983. 80(7): p. 1821-5.
77. Liu, R., et al., *Regulation of CSF1 promoter by the SWI/SNF-like BAF complex*. Cell, 2001. 106(3): p. 309-18.
78. Rich, A. and S. Zhang, *Timeline: Z-DNA: the long road to biological function*. Nat Rev Genet, 2003. 4(7): p. 566-72.
79. Agarwal, A.K., *Transcriptional influence of two poly purine-pyrimidine tracts located in the HSD11B2 (11beta-hydroxysteroid dehydrogenase type 2) gene*. Endocr Res, 2001. 27(1-2): p. 1-9.
80. Naylor, L.H. and E.M. Clark, *d(TG)_n.d(CA)_n sequences upstream of the rat prolactin gene form Z-DNA and inhibit gene transcription*. Nucleic Acids Res, 1990. 18(6): p. 1595-601.
81. Herbert, A. and A. Rich, *The role of binding domains for dsRNA and Z-DNA in the in vivo editing of minimal substrates by ADAR1*. Proc Natl Acad Sci U S A, 2001. 98(21): p. 12132-7.
82. Oh, D.B., Y.G. Kim, and A. Rich, *Z-DNA-binding proteins can act as potent effectors of gene expression in vivo*. Proc Natl Acad Sci U S A, 2002. 99(26): p. 16666-71.
83. Placido, D., et al., *A left-handed RNA double helix bound by the Z alpha domain of the RNA-editing enzyme ADAR1*. Structure, 2007. 15(4): p. 395-404.
84. Teplova, M., et al., *Structural insights into RNA recognition by the alternate-splicing regulator CUG-binding protein 1*. Structure, 2010. 18(10): p. 1364-77.
85. Feng, S., et al., *Alternate rRNA secondary structures as regulators of translation*. Nat Struct Mol Biol, 2011. 18(2): p. 169-76.
86. Polymenidou, M., et al., *Long pre-mRNA depletion and RNA missplicing contribute to neuronal vulnerability from loss of TDP-43*. Nat Neurosci, 2011. 14(4): p. 459-68.
87. Tang, D.C., et al., *Polymorphisms within the angiotensinogen gene (GT-repeat) and the risk of stroke in pediatric patients with sickle cell disease: a case-control study*. Am J Hematol, 2001. 68(3): p. 164-9.
88. Mercado, P.A., et al., *Depletion of TDP 43 overrides the need for exonic and intronic splicing enhancers in the human apoA-II gene*. Nucleic Acids Res, 2005. 33(18): p. 6000-10.
89. Brousseau, T., et al., *Significant impact of the highly informative (CA)_n repeat polymorphism of the APOA-II gene on the plasma APOA-II concentrations and HDL subfractions: The ECTIM study*. Am J Med Genet, 2002. 110(1): p. 19-24.
90. Xu, M., et al., *Genetic variants in the BDNF gene and therapeutic response to risperidone in schizophrenia patients: a pharmacogenetic study*. Eur J Hum Genet, 2010. 18(6): p. 707-12.
91. Hawi, Z., et al., *No linkage or linkage disequilibrium between brain-derived neurotrophic factor (BDNF) dinucleotide repeat polymorphism and schizophrenia in Irish families*. Psychiatry Res, 1998. 81(2): p. 111-6.
92. Muglia, P., et al., *Association between the BDNF gene and schizophrenia*. Mol Psychiatry, 2003. 8(2): p. 146-7.
93. Sasaki, T., et al., *Brain-derived neurotrophic factor gene and schizophrenia in Japanese subjects*. Am J Med Genet, 1997. 74(4): p. 443-4.
94. Virgos, C., et al., *Association study of schizophrenia with polymorphisms at six candidate genes*. Schizophr Res, 2001. 49(1-2): p. 65-71.

95. Krebs, M.O., et al., *Brain derived neurotrophic factor (BDNF) gene variants association with age at onset and therapeutic response in schizophrenia*. Mol Psychiatry, 2000. **5**(5): p. 558-62.
96. Neves-Pereira, M., et al., *The brain-derived neurotrophic factor gene confers susceptibility to bipolar disorder: evidence from a family-based association study*. Am J Hum Genet, 2002. **71**(3): p. 651-5.
97. Gonzalez, A.A., et al., *A polymorphic GT short tandem repeat affecting beta-ENaC mRNA expression is associated with low renin essential hypertension*. Am J Hypertens, 2007. **20**(7): p. 800-6.
98. Nowacka-Zawisza, M., et al., *Dinucleotide repeat polymorphisms of RAD51, BRCA1, BRCA2 gene regions in breast cancer*. Pathol Int, 2008. **58**(5): p. 275-81.
99. Nolan, L.S., et al., *Identification and functional analysis of common sequence variants in the DFNA15 gene, Brn-3c*. Gene, 2007. **400**(1-2): p. 89-97.
100. Omi, K., et al., *Polymorphisms of CD36 in Thai malaria patients*. Southeast Asian J Trop Med Public Health, 2002. **33** Suppl 3: p. 1-4.
101. Omi, K., et al., *CD36 polymorphism is associated with protection from cerebral malaria*. Am J Hum Genet, 2003. **72**(2): p. 364-74.
102. Lipsky, R.H., H. Ikeda, and E.S. Medved, *A dinucleotide repeat in the third intron of CD36*. Hum Mol Genet, 1994. **3**(1): p. 217.
103. Yun, Y.M., et al., *CD36 polymorphism and its relationship with body mass index and coronary artery disease in a Korean population*. Clin Chem Lab Med, 2007. **45**(10): p. 1277-82.
104. Lecompte, S., et al., *Polymorphisms in the CD36/FAT gene are associated with plasma vitamin E concentrations in humans*. Am J Clin Nutr, 2011. **93**(3): p. 644-51.
105. Love-Gregory, L., et al., *Variants in the CD36 gene associate with the metabolic syndrome and high-density lipoprotein cholesterol*. Hum Mol Genet, 2008. **17**(11): p. 1695-704.
106. Ma, X., et al., *A common haplotype at the CD36 locus is associated with high free fatty acid levels and increased cardiovascular risk in Caucasians*. Hum Mol Genet, 2004. **13**(19): p. 2197-205.
107. Sato, M., et al., *A GT dinucleotide repeat polymorphism in intron 1 of the H-cadherin (CDH13) gene*. J Hum Genet, 1998. **43**(4): p. 285-6.
108. Sato, M., et al., *The H-cadherin (CDH13) gene is inactivated in human lung cancer*. Hum Genet, 1998. **103**(1): p. 96-101.
109. Cuppens, H., et al., *Polyvariant mutant cystic fibrosis transmembrane conductance regulator genes. The polymorphic (Tg)m locus explains the partial penetrance of the T5 polymorphism as a disease mutation*. J Clin Invest, 1998. **101**(2): p. 487-96.
110. Groman, J.D., et al., *Variation in a repeat sequence determines whether a common variant of the cystic fibrosis transmembrane conductance regulator gene is pathogenic or benign*. Am J Hum Genet, 2004. **74**(1): p. 176-9.
111. Radpour, R., et al., *Molecular study of (TG)m(T)n polymorphisms in Iranian males with congenital bilateral absence of the vas deferens*. J Androl, 2007. **28**(4): p. 541-7.
112. Huang, Q., W. Ding, and M.X. Wei, *Comparative analysis of common CFTR polymorphisms poly-T, TG-repeats and M470V in a healthy Chinese population*. World J Gastroenterol, 2008. **14**(12): p. 1925-30.
113. Sharma, S., et al., *A novel (TG)n(GA)m repeat polymorphism 254 bp downstream of the mast cell chymase (CMA1) gene is associated with atopic asthma and total serum IgE levels*. J Hum Genet, 2005. **50**(6): p. 276-82.
114. Ma, J., et al., *Association study of a (TG)n dinucleotide repeat at chromosome 15q13.3 and schizophrenia in the Chinese population*. Psychiatry Res, 2008. **159**(1-2): p. 245-9.
115. Cai, Q., et al., *Association of breast cancer risk with a GT dinucleotide repeat polymorphism upstream of the estrogen receptor-alpha gene*. Cancer Res, 2003. **63**(18): p. 5727-30.

116. Boyapati, S.M., et al., *Polymorphisms in ER-alpha gene interact with estrogen receptor status in breast cancer survival*. Clin Cancer Res, 2005. 11(3): p. 1093-8.
117. Curran, J.E., et al., *Association of estrogen receptor and glucocorticoid receptor gene polymorphisms with sporadic breast cancer*. Int J Cancer, 2001. 95(4): p. 271-5.
118. Chen, T.M., et al., *Microsatellite in the 3' untranslated region of human fibroblast growth factor 9 (FGF9) gene exhibits pleiotropic effect on modulating FGF9 protein expression*. Hum Mutat, 2007. 28(1): p. 98.
119. Kim, S.Y., et al., *Dinucleotide repeat polymorphism in Fms-like tyrosine kinase-1 (Flt-1) gene is not associated with preeclampsia*. BMC Med Genet, 2008. 9: p. 68.
120. Iwase, K., et al., *FOXP3/Scurfin gene polymorphism is associated with adult onset type 1 diabetes in Japanese, especially in women and slowly progressive-type patients*. Autoimmunity, 2009. 42(2): p. 159-67.
121. Bassuny, W.M., et al., *A functional polymorphism in the promoter/enhancer region of the FOXP3/Scurfin gene associated with type 1 diabetes*. Immunogenetics, 2003. 55(3): p. 149-56.
122. Nakanishi, K. and Y. Shima, *No contribution of a GT microsatellite polymorphism in the promoter region of the FOXP3 gene to susceptibility to type 1 diabetes in the Japanese population*. Clin Chim Acta, 2007. 384(1-2): p. 171-3.
123. Sanchez, E., et al., *Analysis of a GT microsatellite in the promoter of the foxp3/scurfin gene in autoimmune diseases*. Hum Immunol, 2005. 66(8): p. 869-73.
124. Tang, J., et al., *Significant linkage and association between a functional (GT)n polymorphism in promoter of the N-methyl-D-aspartate receptor subunit gene (GRIN2A) and schizophrenia*. Neurosci Lett, 2006. 409(1): p. 80-2.
125. D'Elia, A.V., et al., *Dinucleotide polymorphisms present in the 5'-flanking region of the human H2 relaxin gene, but not in the corresponding region of the H1 gene*. Horm Metab Res, 2003. 35(1): p. 24-8.
126. Tizzano, E., et al., *Utility of a (GT) dinucleotide repeat in intron 1 of the factor 8 gene for haemophilia A carrier diagnosis*. Haemophilia, 2005. 11(2): p. 142-4.
127. Vencesla, A., et al., *Application of intron 9 and intron 25 dinucleotide repeats of the factor VIII gene for carrier diagnosis in haemophilia A*. Haemophilia, 2008. 14(3): p. 489-93.
128. Arrufat, F.J., et al., *Analysis of the polymorphic (GT)(n) repeat at the dopamine beta-hydroxylase gene in Spanish patients affected by schizophrenia*. Am J Med Genet, 2000. 96(1): p. 88-92.
129. H Lin, G.C., Z Li, N Wang, Y Bai, X Luo, B Yang, Z Wang, *TG repeat length polymorphism in the promoter region of KCNH2 gene controls the transcription of hERG K+ channel*. Abstract from the " Canadian Cardiovascular Congress 2010", 2010.
130. Hadjiyannakis, S., et al., *GT repeat polymorphism in the 5' flanking region of the human growth hormone receptor gene*. Mol Cell Probes, 2001. 15(4): p. 239-42.
131. Koukourakis, M.I., et al., *C2028T polymorphism in exon 12 and dinucleotide repeat polymorphism in intron 13 of the HIF-1alpha gene define HIF-1alpha protein expression in non-small cell lung cancer*. Lung Cancer, 2006. 53(3): p. 257-62.
132. Lin, R., et al., *Association of heme oxygenase-1 gene polymorphisms with essential hypertension and blood pressure in the Chinese Han population*. Genet Test Mol Biomarkers, 2011. 15(1-2): p. 23-8.
133. Wu, M.M., et al., *GT-repeat polymorphism in the heme oxygenase-1 gene promoter and the risk of carotid atherosclerosis related to arsenic exposure*. J Biomed Sci, 2010. 17: p. 70.
134. Bozkaya, O.G., et al., *Prolonged unconjugated hyperbilirubinaemia associated with the haem oxygenase-1 gene promoter polymorphism*. Acta Paediatr, 2010. 99(5): p. 679-83.
135. Lo, S.S., et al., *Heme oxygenase-1 gene promoter polymorphism is associated with risk of gastric adenocarcinoma and lymphovascular tumor invasion*. Ann Surg Oncol, 2007. 14(8): p. 2250-6.

136. Kuesap, J., et al., *Study on association between genetic polymorphisms of haem oxygenase-1, tumour necrosis factor, cadmium exposure and malaria pathogenicity and severity*. Malar J, 2010. **9**: p. 260.
137. Bai, C.H., et al., *Shorter GT repeat polymorphism in the heme oxygenase-1 gene promoter has protective effect on ischemic stroke in dyslipidemia patients*. J Biomed Sci, 2010. **17**: p. 12.
138. Yoshida, K., et al., *Promoter polymorphisms of the interferon-alpha receptor gene and development of Interferon-induced depressive symptoms in patients with chronic hepatitis C: preliminary findings*. Neuropsychobiology, 2005. **52**(2): p. 55-61.
139. Navaglia, F., et al., *Interleukin 12 gene polymorphisms enhance gastric cancer risk in H pylori infected individuals*. J Med Genet, 2005. **42**(6): p. 503-10.
140. Mellick, G.D., et al., *The monoamine oxidase B gene GT repeat polymorphism and Parkinson's disease in a Chinese population*. J Neurol, 2000. **247**(1): p. 52-5.
141. Grimsby, J., et al., *Dinucleotide repeat (TG)₂₃ polymorphism in the MAOB gene*. Nucleic Acids Res, 1992. **20**(4): p. 924.
142. Craddock, N., et al., *No evidence for allelic association between bipolar disorder and monoamine oxidase A gene polymorphisms*. Am J Med Genet, 1995. **60**(4): p. 322-4.
143. Oda, M., et al., *Dinucleotide repeat polymorphisms in the neprilysin gene are not associated with sporadic Alzheimer's disease*. Neurosci Lett, 2002. **320**(1-2): p. 105-7.
144. Choi YS, L.K., Park YG, Kim SH, *Polymorphism Study of Susceptibility Genes Related to Sporadic Alzheimer's Disease*. J Korean Geriatr Soc, 2006. **10**(2):115-124.
145. Texereau, J., et al., *Nitric oxide synthase 1 as a potential modifier gene of decline in lung function in patients with cystic fibrosis*. Thorax, 2004. **59**(2): p. 156-8.
146. Indo, Y., et al., *Structure and organization of the human TRKA gene encoding a high affinity receptor for nerve growth factor*. Jpn J Hum Genet, 1997. **42**(2): p. 343-51.
147. Miura, Y., et al., *Mutation and polymorphism analysis of the TRKA (NTRK1) gene encoding a high-affinity receptor for nerve growth factor in congenital insensitivity to pain with anhidrosis (CIPA) families*. Hum Genet, 2000. **106**(1): p. 116-24.
148. Eskdale, J., et al., *A polymorphic microsatellite marker in the human p55 TNF receptor, CD120a*. Genes Immun, 2000. **1**(3): p. 228-30.
149. McDermott, M.F., et al., *Germline mutations in the extracellular domains of the 55 kDa TNF receptor, TNFR1, define a family of dominantly inherited autoinflammatory syndromes*. Cell, 1999. **97**(1): p. 133-44.
150. Gossen, M., A. Wullrich, and M.W. Kilimann, *Dinucleotide repeat polymorphism within the PHKA1 gene at Xq12-q13*. Hum Genet, 1995. **95**(4): p. 469-70.
151. Sjakste, T., et al., *Identification of an intronic TG repeat polymorphism in the human proteasome core particle PROS-27K gene*. DNA Seq, 2002. **13**(3): p. 139-43.
152. Sjakste, T., et al., *Association of microsatellite polymorphisms of the human 14q13.2 region with type 2 diabetes mellitus in Latvian and Finnish populations*. Ann Hum Genet, 2007. **71**(Pt 6): p. 772-6.
153. Kalis, M., Sjakste, T., Sjakste, N., *Association study between (TG) repeat polymorphism in PSMA6 gene and type II diabetes mellitus in Botnia*. Biologija, 2002. **vol. 2**: p. pp. 12-14.
154. McClatchey, A.I., et al., *Dinucleotide repeat polymorphisms at the SCN4A locus suggest allelic heterogeneity of hyperkalemic periodic paralysis and paramyotonia congenita*. Am J Hum Genet, 1992. **50**(5): p. 896-901.
155. Searle, S. and J.M. Blackwell, *Evidence for a functional repeat polymorphism in the promoter of the human NRAM1 gene that correlates with autoimmune versus infectious disease susceptibility*. J Med Genet, 1999. **36**(4): p. 295-9.
156. O'Brien, B.A., et al., *Association of SLC11A1 promoter polymorphisms with the incidence of autoimmune and inflammatory diseases: a meta-analysis*. J Autoimmun, 2008. **31**(1): p. 42-51.

157. Li, X., et al., *SLC11A1 (NRAMP1) polymorphisms and tuberculosis susceptibility: updated systematic review and meta-analysis*. PLoS One, 2011. **6**(1): p. e15831.
158. Gazouli, M., et al., *Role of functional polymorphisms of NRAMP1 gene for the development of Crohn's disease*. Inflamm Bowel Dis, 2008. **14**(10): p. 1323-30.
159. Gamba, G., *Molecular physiology and pathophysiology of electroneutral cation-chloride cotransporters*. Physiol Rev, 2005. **85**(2): p. 423-93.
160. Zuo Lunli, Z.L., Li Nanfang, *The Relationship between GT Repeat Sequence Polymorphism of Na-K-2Cl Cotransporter NKCC2 SLC12A1 Gene and Essential Hypertension in Kazakhs of Xinjiang*. Abstract from the International Symposium "Salt, Other Minerals and Hypertension 2006", 2006.
161. Gabellini, N., *A polymorphic GT repeat from the human cardiac Na+Ca2+ exchanger intron 2 activates splicing*. Eur J Biochem, 2001. **268**(4): p. 1076-83.
162. Gao, P.S., et al., *Variation in dinucleotide (GT) repeat sequence in the first exon of the STAT6 gene is associated with atopic asthma and differentially regulates the promoter activity in vitro*. J Med Genet, 2004. **41**(7): p. 535-9.
163. Tamura, K., et al., *Novel dinucleotide repeat polymorphism in the first exon of the STAT-6 gene is associated with allergic diseases*. Clin Exp Allergy, 2001. **31**(10): p. 1509-14.
164. Duetsch, G., et al., *STAT6 as an asthma candidate gene: polymorphism-screening, association and haplotype analysis in a Caucasian sib-pair study*. Hum Mol Genet, 2002. **11**(6): p. 613-21.
165. Conrad, C., et al., *Genetic evidence for the involvement of tau in progressive supranuclear palsy*. Ann Neurol, 1997. **41**(2): p. 277-81.
166. Moore, C.E., et al., *Lack of association between Toll-like receptor 2 polymorphisms and susceptibility to severe disease caused by Staphylococcus aureus*. Clin Diagn Lab Immunol, 2004. **11**(6): p. 1194-7.
167. Boraska Jelavic, T., et al., *Microsatellite GT polymorphism in the toll-like receptor 2 is associated with colorectal cancer*. Clin Genet, 2006. **70**(2): p. 156-60.
168. Mokone, G.G., et al., *The guanine-thymine dinucleotide repeat polymorphism within the tenascin-C gene is associated with achilles tendon injuries*. Am J Sports Med, 2005. **33**(7): p. 1016-21.
169. Hough, C., et al., *Influence of a GT repeat element on shear stress responsiveness of the VWF gene promoter*. J Thromb Haemost, 2008. **6**(7): p. 1183-90.
170. Zaahl, M.G., et al., *Expression of the SLC11A1 (NRAMP1) 5'-(GT)_n repeat: opposite effect in the presence of -237C-->T*. Blood Cells Mol Dis, 2004. **33**(1): p. 45-50.
171. Tamanoue, Y., et al., *Polypyrimidine tract-binding protein is required for the repression of gene expression by all-trans retinoic acid*. Dev Growth Differ, 2010. **52**(5): p. 469-79.
172. Williams, S.S., et al., *Large-scale reprogramming of cranial neural crest gene expression by retinoic acid exposure*. Physiol Genomics, 2004. **19**(2): p. 184-97.
173. Buratti, E., et al., *Nuclear factor TDP-43 and SR proteins promote in vitro and in vivo CFTR exon 9 skipping*. EMBO J, 2001. **20**(7): p. 1774-84.
174. Niksic, M., et al., *Functional analysis of cis-acting elements regulating the alternative splicing of human CFTR exon 9*. Hum Mol Genet, 1999. **8**(13): p. 2339-49.
175. Pagani, F., et al., *Splicing factors induce cystic fibrosis transmembrane regulator exon 9 skipping through a nonevolutionary conserved intronic element*. J Biol Chem, 2000. **275**(28): p. 21041-7.
176. Delaney, S.J., et al., *Cystic fibrosis transmembrane conductance regulator splice variants are not conserved and fail to produce chloride channels*. Nat Genet, 1993. **4**(4): p. 426-31.
177. Strong, T.V., et al., *Expression of an abundant alternatively spliced form of the cystic fibrosis transmembrane conductance regulator (CFTR) gene is not associated with a cAMP-activated chloride conductance*. Hum Mol Genet, 1993. **2**(3): p. 225-30.

178. Buratti, E. and F.E. Baralle, *Characterization and functional implications of the RNA binding properties of nuclear factor TDP-43, a novel splicing regulator of CFTR exon 9*. J Biol Chem, 2001. **276**(39): p. 36337-43.
179. Mori, D., et al., *Quantitative analysis of CUG-BP1 binding to RNA repeats*. J Biochem, 2008. **143**(3): p. 377-83.
180. Takahashi, N., et al., *The CUG-binding protein binds specifically to UG dinucleotide repeats in a yeast three-hybrid system*. Biochem Biophys Res Commun, 2000. **277**(2): p. 518-23.
181. Lu, X., N.A. Timchenko, and L.T. Timchenko, *Cardiac elav-type RNA-binding protein (ETR-3) binds to RNA CUG repeats expanded in myotonic dystrophy*. Hum Mol Genet, 1999. **8**(1): p. 53-60.
182. Dujardin, G., et al., *CELF proteins regulate CFTR pre-mRNA splicing: essential role of the divergent domain of ETR-3*. Nucleic Acids Res, 2010. **38**(20): p. 7273-85.
183. Faustino, N.A. and T.A. Cooper, *Identification of putative new splicing targets for ETR-3 using sequences identified by systematic evolution of ligands by exponential enrichment*. Mol Cell Biol, 2005. **25**(3): p. 879-87.
184. Zhang, T., et al., *ApoA-II directs morphogenetic movements of zebrafish embryo by preventing chromosome fusion during nuclear division in yolk syncytial layer*. J Biol Chem, 2011. **286**(11): p. 9514-25.
185. Shelley, C.S. and F.E. Baralle, *Deletion analysis of a unique 3' splice site indicates that alternating guanine and thymine residues represent an efficient splicing signal*. Nucleic Acids Res, 1987. **15**(9): p. 3787-99.
186. Miller, J.H. and D.A. Warrell, *Imported falciparum malaria*. BMJ, 1994. **309**(6963): p. 1233-4.
187. Asch, A.S., et al., *Analysis of CD36 binding domains: ligand specificity controlled by dephosphorylation of an ectodomain*. Science, 1993. **262**(5138): p. 1436-40.
188. Baruch, D.I., et al., *CD36 peptides that block cytoadherence define the CD36 binding region for Plasmodium falciparum-infected erythrocytes*. Blood, 1999. **94**(6): p. 2121-7.
189. McGilvray, I.D., et al., *Nonopsonic monocyte/macrophage phagocytosis of Plasmodium falciparum-parasitized erythrocytes: a role for CD36 in malarial clearance*. Blood, 2000. **96**(9): p. 3231-40.
190. Pain, A., et al., *A non-sense mutation in Cd36 gene is associated with protection from severe malaria*. Lancet, 2001. **357**(9267): p. 1502-3.
191. Aitman, T.J., et al., *Malaria susceptibility and CD36 mutation*. Nature, 2000. **405**(6790): p. 1015-6.
192. Quednau, B.D., D.A. Nicoll, and K.D. Philipson, *Tissue specificity and alternative splicing of the Na⁺/Ca²⁺ exchanger isoforms NCX1, NCX2, and NCX3 in rat*. Am J Physiol, 1997. **272**(4 Pt 1): p. C1250-61.
193. Kofuji, P., et al., *Expression of the Na-Ca exchanger in diverse tissues: a study using the cloned human cardiac Na-Ca exchanger*. Am J Physiol, 1992. **263**(6 Pt 1): p. C1241-9.
194. Capel, B., et al., *Circular transcripts of the testis-determining gene Sry in adult mouse testis*. Cell, 1993. **73**(5): p. 1019-30.
195. Zaphiropoulos, P.G., *Circular RNAs from transcripts of the rat cytochrome P450 2C24 gene: correlation with exon skipping*. Proc Natl Acad Sci U S A, 1996. **93**(13): p. 6536-41.
196. Wang, H.Y., et al., *Structural diversity and functional implications of the eukaryotic TDP gene family*. Genomics, 2004. **83**(1): p. 130-9.
197. Johnson, B.S., et al., *A yeast TDP-43 proteinopathy model: Exploring the molecular determinants of TDP-43 aggregation and cellular toxicity*. Proc Natl Acad Sci U S A, 2008. **105**(17): p. 6439-44.
198. Ayala, Y.M., et al., *Human, Drosophila, and C.elegans TDP43: nucleic acid binding properties and splicing regulatory function*. J Mol Biol, 2005. **348**(3): p. 575-88.

199. Ayala, Y.M., T. Misteli, and F.E. Baralle, *TDP-43 regulates retinoblastoma protein phosphorylation through the repression of cyclin-dependent kinase 6 expression*. Proc Natl Acad Sci U S A, 2008. **105**(10): p. 3785-9.
200. Kuo, P.H., et al., *Structural insights into TDP-43 in nucleic-acid binding and domain interactions*. Nucleic Acids Res, 2009. **37**(6): p. 1799-808.
201. Shiina, Y., et al., *TDP-43 dimerizes in human cells in culture*. Cell Mol Neurobiol, 2010. **30**(4): p. 641-52.
202. Buratti, E., et al., *TDP-43 binds heterogeneous nuclear ribonucleoprotein A/B through its C-terminal tail: an important region for the inhibition of cystic fibrosis transmembrane conductance regulator exon 9 splicing*. J Biol Chem, 2005. **280**(45): p. 37572-84.
203. D'Ambrogio, A., et al., *Functional mapping of the interaction between TDP-43 and hnRNP A2 in vivo*. Nucleic Acids Res, 2009. **37**(12): p. 4116-26.
204. Fuentealba, R.A., et al., *Interaction with polyglutamine aggregates reveals a Q/N-rich domain in TDP-43*. J Biol Chem, 2010. **285**(34): p. 26304-14.
205. Winton, M.J., et al., *Disturbance of nuclear and cytoplasmic TAR DNA-binding protein (TDP-43) induces disease-like redistribution, sequestration, and aggregate formation*. J Biol Chem, 2008. **283**(19): p. 13302-9.
206. Ayala, Y.M., et al., *Structural determinants of the cellular localization and shuttling of TDP-43*. J Cell Sci, 2008. **121**(Pt 22): p. 3778-85.
207. Turner, B.J., et al., *TDP-43 expression in mouse models of amyotrophic lateral sclerosis and spinal muscular atrophy*. BMC Neurosci, 2008. **9**: p. 104.
208. Thorpe, J.R., et al., *Fine structural analysis of the neuronal inclusions of frontotemporal lobar degeneration with TDP-43 proteinopathy*. J Neural Transm, 2008. **115**(12): p. 1661-71.
209. Cmarko, D., et al., *Ultrastructural analysis of transcription and splicing in the cell nucleus after bromo-UTP microinjection*. Mol Biol Cell, 1999. **10**(1): p. 211-23.
210. Lin, S.L., D.C. Chang, and S.Y. Ying, *Isolation and identification of gene-specific microRNAs*. Methods Mol Biol, 2006. **342**: p. 313-20.
211. Casafont, I., et al., *TDP-43 localizes in mRNA transcription and processing sites in mammalian neurons*. J Struct Biol, 2009. **167**(3): p. 235-41.
212. Neumann, M., et al., *Ubiquitinated TDP-43 in frontotemporal lobar degeneration and amyotrophic lateral sclerosis*. Science, 2006. **314**(5796): p. 130-3.
213. Sato, T., et al., *Axonal ligation induces transient redistribution of TDP-43 in brainstem motor neurons*. Neuroscience, 2009. **164**(4): p. 1565-78.
214. Moisse, K., et al., *Cytosolic TDP-43 expression following axotomy is associated with caspase 3 activation in NFL-/- mice: support for a role for TDP-43 in the physiological response to neuronal injury*. Brain Res, 2009. **1296**: p. 176-86.
215. Wang, I.F., et al., *TDP-43, the signature protein of FTL-DU, is a neuronal activity-responsive factor*. J Neurochem, 2008. **105**(3): p. 797-806.
216. Volkering, K., et al., *Tar DNA binding protein of 43 kDa (TDP-43), 14-3-3 proteins and copper/zinc superoxide dismutase (SOD1) interact to modulate NFL mRNA stability. Implications for altered RNA processing in amyotrophic lateral sclerosis (ALS)*. Brain Res, 2009. **1305**: p. 168-82.
217. McDonald, K.K., et al., *TAR DNA-binding protein 43 (TDP-43) regulates stress granule dynamics via differential regulation of G3BP and TIA-1*. Hum Mol Genet, 2011. **20**(7): p. 1400-10.
218. Arai, T., et al., *TDP-43 is a component of ubiquitin-positive tau-negative inclusions in frontotemporal lobar degeneration and amyotrophic lateral sclerosis*. Biochem Biophys Res Commun, 2006. **351**(3): p. 602-11.
219. Lagier-Tourenne, C., M. Polymenidou, and D.W. Cleveland, *TDP-43 and FUS/TLS: emerging roles in RNA processing and neurodegeneration*. Hum Mol Genet, 2010. **19**(R1): p. R46-64.
220. Gitcho, M.A., et al., *VCP mutations causing frontotemporal lobar degeneration disrupt localization of TDP-43 and induce cell death*. J Biol Chem, 2009. **284**(18): p. 12384-98.
221. Kovacs, G.G., et al., *TARDBP variation associated with frontotemporal dementia, supranuclear gaze palsy, and chorea*. Mov Disord, 2009. **24**(12): p. 1843-7.

222. Gitcho, M.A., et al., *TDP-43 A315T mutation in familial motor neuron disease*. Ann Neurol, 2008. **63**(4): p. 535-8.
223. Gitcho, M.A., et al., *TARDBP 3'-UTR variant in autopsy-confirmed frontotemporal lobar degeneration with TDP-43 proteinopathy*. Acta Neuropathol, 2009. **118**(5): p. 633-45.
224. Borroni, B., et al., *Mutation within TARDBP leads to frontotemporal dementia without motor neuron disease*. Hum Mutat, 2009. **30**(11): p. E974-83.
225. Zhang, Y.J., et al., *Aberrant cleavage of TDP-43 enhances aggregation and cellular toxicity*. Proc Natl Acad Sci U S A, 2009. **106**(18): p. 7607-12.
226. Igaz, L.M., et al., *Expression of TDP-43 C-terminal Fragments in Vitro Recapitulates Pathological Features of TDP-43 Proteinopathies*. J Biol Chem, 2009. **284**(13): p. 8516-24.
227. Igaz, L.M., et al., *Enrichment of C-terminal fragments in TAR DNA-binding protein-43 cytoplasmic inclusions in brain but not in spinal cord of frontotemporal lobar degeneration and amyotrophic lateral sclerosis*. Am J Pathol, 2008. **173**(1): p. 182-94.
228. Neumann, M., *Molecular neuropathology of TDP-43 proteinopathies*. Int J Mol Sci, 2009. **10**(1): p. 232-46.
229. Neumann, M., et al., *Phosphorylation of S409/410 of TDP-43 is a consistent feature in all sporadic and familial forms of TDP-43 proteinopathies*. Acta Neuropathol, 2009. **117**(2): p. 137-49.
230. Kawaguchi, Y., et al., *The deacetylase HDAC6 regulates aggresome formation and cell viability in response to misfolded protein stress*. Cell, 2003. **115**(6): p. 727-38.
231. Boyault, C., et al., *HDAC6 controls major cell response pathways to cytotoxic accumulation of protein aggregates*. Genes Dev, 2007. **21**(17): p. 2172-81.
232. Lagier-Tourenne, C. and D.W. Cleveland, *Rethinking ALS: the FUS about TDP-43*. Cell, 2009. **136**(6): p. 1001-4.
233. Chen-Plotkin, A.S., et al., *Variations in the progranulin gene affect global gene expression in frontotemporal lobar degeneration*. Hum Mol Genet, 2008. **17**(10): p. 1349-62.
234. Mishra, M., et al., *Gene expression analysis of frontotemporal lobar degeneration of the motor neuron disease type with ubiquitinated inclusions*. Acta Neuropathol, 2007. **114**(1): p. 81-94.
235. Strong, M.J., et al., *TDP43 is a human low molecular weight neurofilament (hNFL) mRNA-binding protein*. Mol Cell Neurosci, 2007. **35**(2): p. 320-7.
236. Sephton, C.F., et al., *Identification of neuronal RNA targets of TDP-43-containing ribonucleoprotein complexes*. J Biol Chem, 2011. **286**(2): p. 1204-15.
237. Ayala, Y.M., et al., *TDP-43 regulates its mRNA levels through a negative feedback loop*. EMBO J, 2011. **30**(2): p. 277-88.
238. Bose, J.K., et al., *TDP-43 overexpression enhances exon 7 inclusion during the survival of motor neuron pre-mRNA splicing*. J Biol Chem, 2008. **283**(43): p. 28852-9.
239. Dreumont, N., et al., *Antagonistic factors control the unproductive splicing of SC35 terminal intron*. Nucleic Acids Res, 2010. **38**(4): p. 1353-66.
240. Ou, S.H., et al., *Cloning and characterization of a novel cellular protein, TDP-43, that binds to human immunodeficiency virus type 1 TAR DNA sequence motifs*. J Virol, 1995. **69**(6): p. 3584-96.
241. Tollervey, J.R., et al., *Characterizing the RNA targets and position-dependent splicing regulation by TDP-43*. Nat Neurosci, 2011. **14**(4): p. 452-8.
242. Fiesel, F.C., et al., *Knockdown of transactive response DNA-binding protein (TDP-43) downregulates histone deacetylase 6*. EMBO J, 2010. **29**(1): p. 209-21.
243. Chiang, P.M., et al., *Deletion of TDP-43 down-regulates Tbc1d1, a gene linked to obesity, and alters body fat metabolism*. Proc Natl Acad Sci U S A, 2010. **107**(37): p. 16320-4.
244. Kim, S.H., et al., *Amyotrophic lateral sclerosis-associated proteins TDP-43 and FUS/TLS function in a common biochemical complex to co-regulate HDAC6 mRNA*. J Biol Chem, 2010. **285**(44): p. 34097-105.

245. Freibaum, B.D., et al., *Global analysis of TDP-43 interacting proteins reveals strong association with RNA splicing and translation machinery*. J Proteome Res, 2010. **9**(2): p. 1104-20.
246. Ling, S.C., et al., *ALS-associated mutations in TDP-43 increase its stability and promote TDP-43 complexes with FUS/TLS*. Proc Natl Acad Sci U S A, 2010. **107**(30): p. 13318-23.
247. Gregory, R.I., et al., *The Microprocessor complex mediates the genesis of microRNAs*. Nature, 2004. **432**(7014): p. 235-40.
248. Bertolotti, A., et al., *hTAF(II)68, a novel RNA/ssDNA-binding protein with homology to the pro-oncoproteins TLS/FUS and EWS is associated with both TFIID and RNA polymerase II*. EMBO J, 1996. **15**(18): p. 5022-31.
249. Kwiatkowski, T.J., Jr., et al., *Mutations in the FUS/TLS gene on chromosome 16 cause familial amyotrophic lateral sclerosis*. Science, 2009. **323**(5918): p. 1205-8.
250. Vance, C., et al., *Mutations in FUS, an RNA processing protein, cause familial amyotrophic lateral sclerosis type 6*. Science, 2009. **323**(5918): p. 1208-11.
251. Woulfe, J., D.A. Gray, and I.R. Mackenzie, *FUS-immunoreactive intranuclear inclusions in neurodegenerative disease*. Brain Pathol, 2010. **20**(3): p. 589-97.
252. Elden, A.C., et al., *Ataxin-2 intermediate-length polyglutamine expansions are associated with increased risk for ALS*. Nature, 2010. **466**(7310): p. 1069-75.
253. Gilks, N., et al., *Stress granule assembly is mediated by prion-like aggregation of TIA-1*. Mol Biol Cell, 2004. **15**(12): p. 5383-98.
254. Schwab, C., et al., *Colocalization of transactivation-responsive DNA-binding protein 43 and huntingtin in inclusions of Huntington disease*. J Neuropathol Exp Neurol, 2008. **67**(12): p. 1159-65.
255. Young, J.I., et al., *Regulation of RNA splicing by the methylation-dependent transcriptional repressor methyl-CpG binding protein 2*. Proc Natl Acad Sci U S A, 2005. **102**(49): p. 17551-8.
256. Amir, R.E., et al., *Rett syndrome is caused by mutations in X-linked MECP2, encoding methyl-CpG-binding protein 2*. Nat Genet, 1999. **23**(2): p. 185-8.
257. Markovtsov, V., et al., *Cooperative assembly of an hnRNP complex induced by a tissue-specific homolog of polypyrimidine tract binding protein*. Mol Cell Biol, 2000. **20**(20): p. 7463-79.
258. Gromak, N., et al., *Antagonistic regulation of alpha-actinin alternative splicing by CELF proteins and polypyrimidine tract binding protein*. RNA, 2003. **9**(4): p. 443-56.
259. Spellman, R., et al., *Regulation of alternative splicing by PTB and associated factors*. Biochem Soc Trans, 2005. **33**(Pt 3): p. 457-60.
260. Sureau, A., et al., *CELF and PTB proteins modulate the inclusion of the beta-tropomyosin exon 6B during myogenic differentiation*. Exp Cell Res, 2011. **317**(1): p. 94-106.
261. Das, R., et al., *Functional coupling of RNAP II transcription to spliceosome assembly*. Genes Dev, 2006. **20**(9): p. 1100-9.
262. Wang, I.F., N.M. Reddy, and C.K. Shen, *Higher order arrangement of the eukaryotic nuclear bodies*. Proc Natl Acad Sci U S A, 2002. **99**(21): p. 13583-8.
263. Bannwarth, S. and A. Gagnon, *HIV-1 TAR RNA: the target of molecular interactions between the virus and its host*. Curr HIV Res, 2005. **3**(1): p. 61-71.
264. Chande, A.G., M. Baba, and R. Mukhopadhyaya, *A single step assay for rapid evaluation of inhibitors targeting HIV-1 Tat mediated LTR transactivation*. AIDS Res Hum Retroviruses, 2011.
265. Gerasimova, T.I., K. Byrd, and V.G. Corces, *A chromatin insulator determines the nuclear localization of DNA*. Mol Cell, 2000. **6**(5): p. 1025-35.
266. Abhyankar, M.M., C. Urekar, and P.P. Reddi, *A novel CpG-free vertebrate insulator silences the testis-specific SP-10 gene in somatic tissues: role for TDP-43 in insulator function*. J Biol Chem, 2007. **282**(50): p. 36143-54.
267. Lalmansingh, A.S., C.J. Urekar, and P.P. Reddi, *TDP-43 Is a Transcriptional Repressor: THE TESTIS-SPECIFIC MOUSE acrv1 GENE IS A TDP-43 TARGET IN VIVO*. J Biol Chem, 2011. **286**(13): p. 10970-82.

268. Acharya, K.K., et al., *cis-requirement for the maintenance of round spermatid-specific transcription*. Dev Biol, 2006. **295**(2): p. 781-90.
269. Reddi, P.P., et al., *Role of an insulator in testis-specific gene transcription*. Ann N Y Acad Sci, 2007. **1120**: p. 95-103.
270. Iguchi, Y., et al., *TDP-43 depletion induces neuronal cell damage through dysregulation of Rho family GTPases*. J Biol Chem, 2009. **284**(33): p. 22059-66.
271. Pandey, U.B., et al., *HDAC6 rescues neurodegeneration and provides an essential link between autophagy and the UPS*. Nature, 2007. **447**(7146): p. 859-63.
272. Ayala, Y.M., F. Pagani, and F.E. Baralle, *TDP43 depletion rescues aberrant CFTR exon 9 skipping*. FEBS Lett, 2006. **580**(5): p. 1339-44.
273. Pagani, F., et al., *New type of disease causing mutations: the example of the composite exonic regulatory elements of splicing in CFTR exon 12*. Hum Mol Genet, 2003. **12**(10): p. 1111-20.
274. Pagani, F., et al., *Promoter architecture modulates CFTR exon 9 skipping*. J Biol Chem, 2003. **278**(3): p. 1511-7.
275. Buratti, E., et al., *RNA structure is a key regulatory element in pathological ATM and CFTR pseudoexon inclusion events*. Nucleic Acids Res, 2007. **35**(13): p. 4369-83.
276. Buratti, E., et al., *SR protein-mediated inhibition of CFTR exon 9 inclusion: molecular characterization of the intronic splicing silencer*. Nucleic Acids Res, 2007. **35**(13): p. 4359-68.
277. Buratti, E., et al., *Nuclear factor TDP-43 binds to the polymorphic TG repeats in CFTR intron 8 and causes skipping of exon 9: a functional link with disease penetrance*. Am J Hum Genet, 2004. **74**(6): p. 1322-5.
278. Zuccato, E., et al., *An intronic polypyrimidine-rich element downstream of the donor site modulates cystic fibrosis transmembrane conductance regulator exon 9 alternative splicing*. J Biol Chem, 2004. **279**(17): p. 16980-8.
279. Arrisi-Mercado, P., et al., *An exonic splicing enhancer offsets the atypical GU-rich 3' splice site of human apolipoprotein A-II exon 3*. J Biol Chem, 2004. **279**(38): p. 39331-9.
280. Lorson, C.L. and E.J. Androphy, *An exonic enhancer is required for inclusion of an essential exon in the SMA-determining gene SMN*. Hum Mol Genet, 2000. **9**(2): p. 259-65.
281. Monani, U.R., et al., *A single nucleotide difference that alters splicing patterns distinguishes the SMA gene SMN1 from the copy gene SMN2*. Hum Mol Genet, 1999. **8**(7): p. 1177-83.
282. Lorson, C.L., et al., *A single nucleotide in the SMN gene regulates splicing and is responsible for spinal muscular atrophy*. Proc Natl Acad Sci U S A, 1999. **96**(11): p. 6307-11.
283. Kashima, T., et al., *hnRNP A1 functions with specificity in repression of SMN2 exon 7 splicing*. Hum Mol Genet, 2007. **16**(24): p. 3149-59.
284. Vezain, M., et al., *A rare SMN2 variant in a previously unrecognized composite splicing regulatory element induces exon 7 inclusion and reduces the clinical severity of spinal muscular atrophy*. Hum Mutat, 2010. **31**(1): p. E1110-25.
285. Kashima, T., N. Rao, and J.L. Manley, *An intronic element contributes to splicing repression in spinal muscular atrophy*. Proc Natl Acad Sci U S A, 2007. **104**(9): p. 3426-31.
286. Singh, N.K., et al., *Splicing of a critical exon of human Survival Motor Neuron is regulated by a unique silencer element located in the last intron*. Mol Cell Biol, 2006. **26**(4): p. 1333-46.
287. Hofmann, Y., et al., *Htra2-beta 1 stimulates an exonic splicing enhancer and can restore full-length SMN expression to survival motor neuron 2 (SMN2)*. Proc Natl Acad Sci U S A, 2000. **97**(17): p. 9618-23.
288. Hofmann, Y. and B. Wirth, *hnRNP-G promotes exon 7 inclusion of survival motor neuron (SMN) via direct interaction with Htra2-beta1*. Hum Mol Genet, 2002. **11**(17): p. 2037-49.

289. Young, P.J., et al., *SRp30c-dependent stimulation of survival motor neuron (SMN) exon 7 inclusion is facilitated by a direct interaction with hTra2 beta 1*. Hum Mol Genet, 2002. **11**(5): p. 577-87.
290. Wollerton, M.C., et al., *Autoregulation of polypyrimidine tract binding protein by alternative splicing leading to nonsense-mediated decay*. Mol Cell, 2004. **13**(1): p. 91-100.
291. Rossbach, O., et al., *Auto- and cross-regulation of the hnRNP L proteins by alternative splicing*. Mol Cell Biol, 2009. **29**(6): p. 1442-51.
292. Sun, S., et al., *SF2/ASF autoregulation involves multiple layers of post-transcriptional and translational control*. Nat Struct Mol Biol, 2010. **17**(3): p. 306-12.
293. Lareau, L.F., et al., *Unproductive splicing of SR genes associated with highly conserved and ultraconserved DNA elements*. Nature, 2007. **446**(7138): p. 926-9.
294. Ni, J.Z., et al., *Ultraconserved elements are associated with homeostatic control of splicing regulators by alternative splicing and nonsense-mediated decay*. Genes Dev, 2007. **21**(6): p. 708-18.
295. Saltzman, A.L., et al., *Regulation of multiple core spliceosomal proteins by alternative splicing-coupled nonsense-mediated mRNA decay*. Mol Cell Biol, 2008. **28**(13): p. 4320-30.
296. Sureau, A., et al., *SC35 autoregulates its expression by promoting splicing events that destabilize its mRNAs*. EMBO J, 2001. **20**(7): p. 1785-96.
297. Wegorzewska, I., et al., *TDP-43 mutant transgenic mice develop features of ALS and frontotemporal lobar degeneration*. Proc Natl Acad Sci U S A, 2009. **106**(44): p. 18809-14.
298. Xu, Y.F., et al., *Wild-type human TDP-43 expression causes TDP-43 phosphorylation, mitochondrial aggregation, motor deficits, and early mortality in transgenic mice*. J Neurosci, 2010. **30**(32): p. 10851-9.
299. Fukuda, T., et al., *DEAD-box RNA helicase subunits of the Drosha complex are required for processing of rRNA and a subset of microRNAs*. Nat Cell Biol, 2007. **9**(5): p. 604-11.
300. Buratti, E., et al., *Nuclear factor TDP-43 can affect selected microRNA levels*. FEBS J, 2010. **277**(10): p. 2268-81.
301. Legesse-Miller, A., et al., *let-7 Overexpression leads to an increased fraction of cells in G2/M, direct down-regulation of Cdc34, and stabilization of Wee1 kinase in primary fibroblasts*. J Biol Chem, 2009. **284**(11): p. 6605-9.
302. Hebert, S.S. and B. De Strooper, *Molecular biology. miRNAs in neurodegeneration*. Science, 2007. **317**(5842): p. 1179-80.
303. Colombrita, C., et al., *TDP-43 is recruited to stress granules in conditions of oxidative insult*. J Neurochem, 2009. **111**(4): p. 1051-61.
304. Tongiorgi, E., et al., *Brain-derived neurotrophic factor mRNA and protein are targeted to discrete dendritic laminae by events that trigger epileptogenesis*. J Neurosci, 2004. **24**(30): p. 6842-52.
305. Tiruchinapalli, D.M., et al., *Activity-dependent trafficking and dynamic localization of zipcode binding protein 1 and beta-actin mRNA in dendrites and spines of hippocampal neurons*. J Neurosci, 2003. **23**(8): p. 3251-61.
306. Steward, O. and P.F. Worley, *Selective targeting of newly synthesized Arc mRNA to active synapses requires NMDA receptor activation*. Neuron, 2001. **30**(1): p. 227-40.
307. Rook, M.S., M. Lu, and K.S. Kosik, *CaMKIIalpha 3' untranslated region-directed mRNA translocation in living neurons: visualization by GFP linkage*. J Neurosci, 2000. **20**(17): p. 6385-93.
308. Knowles, R.B., et al., *Translocation of RNA granules in living neurons*. J Neurosci, 1996. **16**(24): p. 7812-20.
309. Ashraf, S.I., et al., *Synaptic protein synthesis associated with memory is regulated by the RISC pathway in Drosophila*. Cell, 2006. **124**(1): p. 191-205.
310. Schratt, G.M., et al., *A brain-specific microRNA regulates dendritic spine development*. Nature, 2006. **439**(7074): p. 283-9.

311. Kress, T.L., Y.J. Yoon, and K.L. Mowry, *Nuclear RNP complex assembly initiates cytoplasmic RNA localization*. J Cell Biol, 2004. **165**(2): p. 203-11.
312. Besse, F. and A. Ephrussi, *Translational control of localized mRNAs: restricting protein synthesis in space and time*. Nat Rev Mol Cell Biol, 2008. **9**(12): p. 971-80.
313. Martin, K.C. and R.S. Zukin, *RNA trafficking and local protein synthesis in dendrites: an overview*. J Neurosci, 2006. **26**(27): p. 7131-4.
314. Sutton, M.A. and E.M. Schuman, *Dendritic protein synthesis, synaptic plasticity, and memory*. Cell, 2006. **127**(1): p. 49-58.
315. Makeyev, A.V. and S.A. Liebhaber, *The poly(C)-binding proteins: a multiplicity of functions and a search for mechanisms*. RNA, 2002. **8**(3): p. 265-78.
316. Thyagarajan, A. and B.G. Szaro, *Phylogenetically conserved binding of specific K homology domain proteins to the 3'-untranslated region of the vertebrate middle neurofilament mRNA*. J Biol Chem, 2004. **279**(48): p. 49680-8.
317. Liu, J., et al., *A role for the P-body component GW182 in microRNA function*. Nat Cell Biol, 2005. **7**(12): p. 1261-6.
318. Anderson, P. and N. Kedersha, *Stress granules: the Tao of RNA triage*. Trends Biochem Sci, 2008. **33**(3): p. 141-50.
319. Dewey, C.M., et al., *TDP-43 is directed to stress granules by sorbitol, a novel physiological osmotic and oxidative stressor*. Mol Cell Biol, 2011. **31**(5): p. 1098-108.
320. Quaresma, A.J., et al., *Human hnRNP Q re-localizes to cytoplasmic granules upon PMA, thapsigargin, arsenite and heat-shock treatments*. Exp Cell Res, 2009. **315**(6): p. 968-80.
321. Deirdre, A., J. Scadden, and C.W. Smith, *Interactions between the terminal bases of mammalian introns are retained in inosine-containing pre-mRNAs*. EMBO J, 1995. **14**(13): p. 3236-46.
322. Baralle, M., et al., *Identification of a mutation that perturbs NF1 agene splicing using genomic DNA samples and a minigene assay*. J Med Genet, 2003. **40**(3): p. 220-2.
323. Stamm, S., et al., *ASD: a bioinformatics resource on alternative splicing*. Nucleic Acids Res, 2006. **34**(Database issue): p. D46-55.
324. Thanaraj, T.A., et al., *ASD: the Alternative Splicing Database*. Nucleic Acids Res, 2004. **32**(Database issue): p. D64-9.
325. Gayther, S.A., et al., *The contribution of germline BRCA1 and BRCA2 mutations to familial ovarian cancer: no evidence for other ovarian cancer-susceptibility genes*. Am J Hum Genet, 1999. **65**(4): p. 1021-9.
326. Claes, K., et al., *Differentiating pathogenic mutations from polymorphic alterations in the splice sites of BRCA1 and BRCA2*. Genes Chromosomes Cancer, 2003. **37**(3): p. 314-20.
327. Foretova, L., et al., *BRCA1 and BRCA2 mutations in women with familial or early-onset breast/ovarian cancer in the Czech Republic*. Hum Mutat, 2004. **23**(4): p. 397-8.
328. Machackova, E., et al., *Spectrum and characterisation of BRCA1 and BRCA2 deleterious mutations in high-risk Czech patients with breast and/or ovarian cancer*. BMC Cancer, 2008. **8**: p. 140.
329. Staley, J.P. and C. Guthrie, *Mechanical devices of the spliceosome: motors, clocks, springs, and things*. Cell, 1998. **92**(3): p. 315-26.
330. Wahl, M.C., C.L. Will, and R. Luhrmann, *The spliceosome: design principles of a dynamic RNP machine*. Cell, 2009. **136**(4): p. 701-18.
331. Reed, R., *Protein composition of mammalian spliceosomes assembled in vitro*. Proc Natl Acad Sci U S A, 1990. **87**(20): p. 8031-5.
332. Konarska, M.M. and P.A. Sharp, *Interactions between small nuclear ribonucleoprotein particles in formation of spliceosomes*. Cell, 1987. **49**(6): p. 763-74.
333. Michaud, S. and R. Reed, *A functional association between the 5' and 3' splice site is established in the earliest prespliceosome complex (E) in mammals*. Genes Dev, 1993. **7**(6): p. 1008-20.

334. Konforti, B.B., M.J. Koziolkiewicz, and M.M. Konarska, *Disruption of base pairing between the 5' splice site and the 5' end of U1 snRNA is required for spliceosome assembly*. Cell, 1993. **75**(5): p. 863-73.
335. Lamond, A.I., M.M. Konarska, and P.A. Sharp, *A mutational analysis of spliceosome assembly: evidence for splice site collaboration during spliceosome formation*. Genes Dev, 1987. **1**(6): p. 532-43.
336. Boehringer, D., et al., *Three-dimensional structure of a pre-catalytic human spliceosomal complex B*. Nat Struct Mol Biol, 2004. **11**(5): p. 463-8.
337. Reed, R., *Mechanisms of fidelity in pre-mRNA splicing*. Curr Opin Cell Biol, 2000. **12**(3): p. 340-5.
338. Balz, V., et al., *Analysis of BRCA1, TP53, and TSG101 germline mutations in German breast and/or ovarian cancer families*. Cancer Genet Cytogenet, 2002. **138**(2): p. 120-7.
339. Smith, T.M., et al., *Complete genomic sequence and analysis of 117 kb of human DNA containing the gene BRCA1*. Genome Res, 1996. **6**(11): p. 1029-49.
340. Szanto, A., et al., *Retinoid X receptors: X-ploring their (patho)physiological functions*. Cell Death Differ, 2004. **11 Suppl 2**: p. S126-43.
341. Maden, M., *Retinoic acid in the development, regeneration and maintenance of the nervous system*. Nat Rev Neurosci, 2007. **8**(10): p. 755-65.
342. Liu, M., et al., *Expression of microsomal epoxide hydrolase is elevated in Alzheimer's hippocampus and induced by exogenous beta-amyloid and trimethyltin*. Eur J Neurosci, 2006. **23**(8): p. 2027-34.
343. Friedrich, B., et al., *The serine/threonine kinases SGK2 and SGK3 are potent stimulators of the epithelial Na⁺ channel alpha,beta,gamma-ENaC*. Pflugers Arch, 2003. **445**(6): p. 693-6.
344. Pao, A.C., et al., *Expression and role of serum and glucocorticoid-regulated kinase 2 in the regulation of Na⁺/H⁺ exchanger 3 in the mammalian kidney*. Am J Physiol Renal Physiol, 2010. **299**(6): p. F1496-506.
345. Hatin, I., et al., *Molecular dissection of translation termination mechanism identifies two new critical regions in eRF1*. Nucleic Acids Res, 2009. **37**(6): p. 1789-98.
346. Alkalaeva, E.Z., et al., *In vitro reconstitution of eukaryotic translation reveals cooperativity between release factors eRF1 and eRF3*. Cell, 2006. **125**(6): p. 1125-36.
347. Inge-Vechtomov, S., G. Zhouravleva, and M. Philippe, *Eukaryotic release factors (eRFs) history*. Biol Cell, 2003. **95**(3-4): p. 195-209.
348. Cartegni, L., S.L. Chew, and A.R. Krainer, *Listening to silence and understanding nonsense: exonic mutations that affect splicing*. Nat Rev Genet, 2002. **3**(4): p. 285-98.
349. Pagani, F. and F.E. Baralle, *Genomic variants in exons and introns: identifying the splicing spoilers*. Nat Rev Genet, 2004. **5**(5): p. 389-96.
350. Huang, J.K., et al., *Retinoid X receptor gamma signaling accelerates CNS remyelination*. Nat Neurosci, 2011. **14**(1): p. 45-53.
351. Latasa, M.J. and J.M. Cosgaya, *Regulation of retinoid receptors by retinoic acid and axonal contact in Schwann cells*. PLoS One, 2011. **6**(2): p. e17023.
352. Kuroyanagi, H., *Fox-1 family of RNA-binding proteins*. Cell Mol Life Sci, 2009. **66**(24): p. 3895-907.
353. Llorian, M., et al., *Position-dependent alternative splicing activity revealed by global profiling of alternative splicing events regulated by PTB*. Nat Struct Mol Biol, 2010. **17**(9): p. 1114-23.
354. Witten, J.T. and J. Ule, *Understanding splicing regulation through RNA splicing maps*. Trends Genet, 2011. **27**(3): p. 89-97.
355. Baralle, M., et al., *NF1 mRNA biogenesis: effect of the genomic milieu in splicing regulation of the NF1 exon 37 region*. FEBS letters, 2006. **580**(18): p. 4449-56.

Appendices

Table 4. Oligonucleotides List. List of the primers used for PCR reactions. The DNA oligonucleotides were purchased from IDT (integrated DNA Technologies); the RNA oligonucleotides from Sigma-Aldrich.

OLIGONUCLEOTIDES FOR REVERSE TRANSCRIPTION

ALPHA 2-3	CAA CTT CAA GCT CCT AAG CCA CTG C
BRA	TAG GAT CCG GTC ACC AGG AAG TTG GTT AAA TCA
PY7 EXON 2 FW	GAA TAC AAG CTT GTC GAG GAG GAC A
PY7 EXON 3 REV	AGA CCG GAA TTC GGA TCC TCT AGA G

OLIGONUCLEOTIDES FOR ENDOGENOUS RNAs AMPLIFICATION

ACSS2 FW	GGG TTT CCC AGT AAG ATG C
ACSS2 REV	CAA AAC ACT GGT GGC ACC A
BRCA1 FW	CTG CTT GTG AAT TTT CTG
BRCA1 REV	CTT TTT CTTG ATG TGC TT
CDC25A EXON 6 FW	GAA GCC TTT GAG TTT AAG
CDC25A EXON 8 REV	CTC ACA AGG TTT GTA GTT
COL4A6 EX24 FW	AAT CAC CCT GCC CTG TAT TA
COL4A6 EX2627 REV	TTC CCG TGC ACA CCC TTG AGT CCT GG
EPHX1 FW	GGC TGG ACA TCC ACT TCA T
EPHX1 REV	CTA CGG TGT CAG GCT TGG T
ETF1 EXON 8 end REV	CTT CTG TGC CTT GGC AAT GAA GA
ETF1 EXON 5	GCT CTT ACA GCA CTA CTT TCA GAT G
ETF1 EXON 6 FW	GTA GAG GAG GTC AGT CAG CCT TG
ETF1 EXON 6 REV	CTG ATC AAA CAT ATC AGA TTG ACT T
ETF1 EXON 7REV	TTA ATT TCT TCT CTT GAA TGA ATT TCA CGT TGG AG
ETF1 EXON 8 short REV	ACA CGG GCA AGT ACT GTT TTG GCG TTG AAG ATA
ETF1 EXON 9 REV	CTC TTT GTC TGT GAA ATG AGA TTT A
HMG1 EXON 4 FW	AAA CCT CCT GCA AAA GTG G
HMG1 EXON 6-281 REV	TAT CCC ACA CTA TTT TCT G
MAP3K19 FW	TTT TGC TAC TTG AGA AGA T
MAP3K19 REV	CTT TTC CTT TGT TCT CAA C
MYH7 FW	ATC CTG GCC GAG TGG AAG
MYH7 REV	CTT CAA CAA GCT CTG GAG
N4BPL2L EXON 4 FW	AGA AAG AGT TTT GAA GAA A
N4BPL2L FW	GTT TGA TAG AAA GCT GAA AAG
N4BPL2L REV	TCT GTT GAG TTT TTT TAT TAT TC
N4BPL2L4 EXON 6 REV	AGT TCA AGG AAT TTT CAT T
NTRK ENDOG FW	AAC ACC TTT TGG GGT CTC GGT GGC TGT
NTRK ENDOG REV	CGC CTC CTT CAG TGC CTT GAC AGC CAC
NTRK1 bis FW	GTC TCG GTG GCT GTG GGC C
NTRK1 bis REV	CTT GAC AGC CAC CAG CAT C

ORC5L FW	GCT GAC AGA AAT GTG ACT G
ORC5L REV	CTT CTC CTT TAA CCA CGG G
PPIA FW	CGT TTT GCA GAC GCC ACC
PPIA REV	CAC TCA GTC TTG GCA GTG
PTPRU FW	CGC GGG CAC CGG CCG CAC A
PTPRU REV	GTC AGT CAG GGC TGC ATT A
RBCK1 FW	AGG AAA TGG CCC TGA GCC
RBCK1 REV	CAT GTC CTT GAG AGA CGC
RGL EXON 10 FW	GAG GTG AAG CTG CAG GTC
RGL EXON 12 REV	CTT CCT CCG CTT GTC AAA
RGL LONGER FW	TCT GTC CAC TGC GAG ATG ACC
RXRG ENDOG EX56 FW	TAT GGA GAA CTC GAC AAA TGA CCC TGT T
RXRG ENDOG EX89	ACA GGC CCT TGG CAT CTG GGT TAA AGA G
SCAP2 FW	TTT ACA GCA GCT TCT CCC A
SCAP2 REV	CTG AGG TGT GAT GGA CAC T
SGK2 EXON 7 FW	CTC TTC TTC CAC CTG CAG
SGK2 EXON 9 REV	GCT GCA GGT GGA AGA AGA GA
SGK2 EXON 6 FW	CAG AGC CAC ATC ATG GCA GAG CGC A
SGK2 EXON 9/10 REV	TCA GGT GCC AAG TAC TCA GGG GTA CCA
SMN EXON 6 FW	CCA GGT CTA AAA TTC AAT G
SMN EXON 8 REV	TTA ATT TAA GGA ATG TGA GC
T3D FW	TGA GCC CCT TCA AGA TAC
T3D REV	CTG ATA GAC CTG GTC ATT
TG EXON 11 REV	TAG ACT GAG CCG CCA GGC G
TG EXON 9 FW	CTG AAG GCC AAT CTT GTG C

OLIGONUCLEOTIDES FOR MINIGENES CONSTRUCTS

BRCA 5'SS pY7 FW	CAA GTC AGG TAA AAA GCG TGT GTG TGT GTG CAC ATG CGT GTG TGT GGT GTC CTT TGC ATT CAA AGC CGC GC
BRCA 5'SS pY7 REV	ATA TGC GCG GGT TGA ATG CAA AGG ACA CCA CAC ACA CGC ATG TGC ACA CAC ACA CAC GCT TTT TAC CTG A
BRCA 5'SS MUT pY7 FW	CAA GTC AAG TAA AAA GCG TGT GTG TGT GTG CAC ATG CGT GTG TGT GGT GTC CTT TGC ATT CAA AGC CGC GC
BRCA 5'SS MUT pY7 REV	ATA TGC GCG GGT TGA ATG CAA AGG ACA CCA CAC ACA CGC ATG TGC ACA CAC ACA CAC GCT TTT TAC TTG A
COL4A6 FW clone	GGA ATT CCA TAT GGA ATT CCC AGT AGT ATT TGG GGA TCC ACA
COL4A6 REV clone	GGA ATT CCA TAT GGA ATT CCC TTT AGT GAT CTC GGA CAA GTC
EPHX1 EXON 5 REV clone	GGA ATT CCA TAT GGA ATT CCC TTC TGG GGA GTT CTG CTC A
EPHX1 FW EXON 5 clone	GGA ATT CCA TAT GGA ATT CCC CCT CGC CAA GGG TGG GC
ETF FILL IN FW	TTA TTC TGT GAA ATG TTT TCC CAA ACA ACC AAA CAA CCA AAC
ETF FILL IN REV	ATA CAT ACA ATC ACG TAC ATT TGG TTG TTT GGT TGT TTG GTT
ETF FILL IN TG NON REP FW	TTA TTC TGT GAA ATG TTT TCG GTT TGT TGG TTT GTT GGT TTG
ETF FILL IN TG NON REP REV	ATA CAT ACA ATC ACG TAC ATA ACC AAC AAA CCA ACA AAC CAA
ΔTG ETF1 FW	GTG AAA TGT TTT CAT GTA CGT GAT TGT ATG TAT GAG AC

ΔTG ETF1 REV	ACA ATC ACG TAC ATG AAA ACA TTT CAC AGA ATA AGT AG
ETF1 EXON 7 FW clone	GGA ATT CCA TAT GGA ATT CCC CTA CTA CCT TAT TTG GCA
ETF1 EXON 7 REV clone	GGA ATT CCA TAT GGA ATT CCC GAT GTT CAA GAT ATA GTC CTG GT
NTRK clone-nde FW	GGA ATT CCA TAT GGA ATT CCC GAG GCG GCA GCT GCT AAT TG
NTRK clone-nde REV	GGA ATT CCA TAT GGA ATT CCC TAA AAT TCA ACC TGC ACC CCT T
NTRK EXON 11 FW clone	AAT TCG GGG TAC CCC GGA GGC GGC AGC TGC TAA TTG G
NTRK EXON 11 REV clone	AAT TCG GGG TAC CCC GGC ATG GGG GCT AAA ATT CAA CC
RXGR clone-nde FW	GGA ATT CCA TAT GGA ATT CCC TAA ATG TTC AAC CCA CA
RXRG clone-nde REV	GGA ATT CCA TAT GGA ATT CCC GAG AAT AGG TTC TCC CAA CT
RXRG EXON 7 FW clone	AAT TCG GGG TAC CCC GTA AAT GTT CAA CCC ACA
RXRG EXON 7 REV clone	AAT TCG GGG TAC CCC GGA GAA TAG GTT CTC CCA ACT G
SGK2 FW clone	GGA ATT CCA TAT GGA ATT CCC TGC TGA CTC TTT GGA CA
SGK2 REV clone	GGA ATT CCA TAT GGA ATT CCC TAT GCT CAA ATG CCT CCT CT
TG 12-18NT FW	CAA GGT ATG TAT CAA GCT TAC ATG TGT GTG TGT GTG TGT GTG TGT GAG ACA GCT TTA AGG AGA CCA ATA GAA CCC GCG C
TG 12-18NT REV	ATA TGC GCG GGT TCT ATT GTT CTC CTT AAA GCT GTC TCA CAC ACA CAC ACA CAC ACA CAC ATG TAA GCT TGA TAC ATA C
TG 12-8NT FW	CAA GGT ATG TAT TGT GTG TGT GTG TGT GTG TGT GTG CAA GCT TAC AAG ACA GCT TTA AGG AGA CCA ATA GAA CCC GCG C
TG 12-8NT REV	ATA TGC GCG GGT TCT ATT GTT CTC CTT AAA GCT GTC TTG TAA GCT TGC ACA CAC ACA CAA TAC ATAC
TG 6-18 FW	CAA GGT ATG TAT CAA GCT TAC ATG TGT GTG TGT GAG ACA GCT TTA AGG AGA CCA ATA GAA CCC GCG C
TG 6-18 REV	ATA TGC GCG GGT TCT ATT GTT CTC CTT AAA GCT GTC TCA CAC ACA CAC ATG TAA GCT TGA TAC ATA C
TG 6-8 FW	CAA GGT ATG TAT TGT GTG TGT GTG CAA GCT TAC AAG ACA GCT TTA AGG AGA CCA ATA GAA CCC GCG C
TG 6-8 REV	ATA TGC GCG GGT TCT ATT GTT CTC CTT AAA GCT GTC TTG TAA GCT TGC ACA CAC ACA CAC ACA CAC ACA CAA TAC ATAC
pY7 Kpn	TAG AGG ATC TGG CTA GTA GCG ATG ACC CTG C
pY7 Xba	TTC GGA TCC TCT AGA G TAG AGG ATC TGG CTA G

OLIGONUCLEOTIDES FOR COUPLED SPLICING-TRANSCRIPTION

CMV FW	TGG AGG TCG CTG AGT AGT GC
BGH REV	TAG AAG GCA CAG TCG AGG

RNA OLIGONUCLEOTIDES

(GU)₆	GUG UGU GUG UGU
(GU)₂A₆(GU)₂	GUG UAA AAA AGU GU

Figures and Tables Legend:

Figure 1. Elements required for proper exon definition.....	3
Figure 2. Tandem repeats characteristics.....	6
Figure 2B. Schematic illustration of TDP-43 domains.....	31
Figure 3. TDP-43 involvement in 3'SS definition.....	52
Figure 4. Examples of TDP-43 involvement in splicing regulation.....	55
Figure 5. Schematic representation of the reporter minigenes used in this study.....	65
Figure 6. Schematic representation of the constructs generated in this study.....	69
Figure 7. Schematic representation of the TDP-43 constructs used in this study.....	72
Figure 8. Insertion of TG6 and 12 at 8 nucleotides from the 5' end of exon 2 does not affect exon 2 processing.....	83
Figure 9. Insertion of TG6 and 12 at 18 nucleotides from the 5' end of exon 2 does not affect exon 2 processing.....	85
Figure 10. Insertion of TG6 and 12 at 18 and 8 nucleotides from the 5' end of exon 2 does not affect exon 2 processing.....	87
Figure 11. BRCA1 5'SS minigene constructs (WT and MUT).....	92
Figure 12. The mutation -1G>A weakens the BRCA1 5'SS and affects exon processing.....	93
Figure 13. BRCA1 5'SS MUT presents delay in exon processing.....	95
Figure 14. The mutation -1G>A leads to exon 12 skipping in constructs carrying the whole length exon 12.....	97
Figure 15. RNA competing oligo GU6 affects exon processing.....	101
Figure 16. RNA competing oligo GU6 affects spliceosome assembly.....	104
Figure 17. Splicing analysis of BRCA1 5'SS WT and MUT in different reporter minigene systems.....	109
Figure 18. Splicing analysis of BRCA1 exon 12 WT and MUT in different reporter minigene systems.....	112
Figure 19. TDP-43 favors BRCA1 exon 12 MUT inclusion in the final mRNA product.....	115
Figure 20. TDP-43 does not affect BRCA1 exon 12 WT inclusion.....	116
Figure 21. RXRG exon 7 constitutive splicing and UG element conservation among mammals.	125
Figure 22. TDP-43 overexpression inhibits RXRG exon 7 inclusion.....	126
Figure 23. TDP-43 levels do not affect exon processing in EPHX1 exon 5, NTRK1 exon 12 and SGK2 exons 7, 8.....	130
Figure 24. ETF1 exon 7 alternative processing and UG element conservation among mammals.	133

Figure 25. TDP-43 overexpression inhibits ETF1 exon 7 inclusion.....	134
Figure 26. Figure 26. The UG repeated element enhances ETF1 exon 7 inclusion...	136
Figure 27. TDP-43 silencing does not affect endogenous exon processing in the exons analyzed so far.....	139
Figure 28. TDP-43 silencing and overexpression do not affect endogenous ETF1 exon 7 processing.....	141
Figure 29. ESE and ISE sequences ensuring RXRG exon 7 inclusion.....	175
Figure 30. ESE and ISE sequences ensuring ETF1 exon 7 and 8 inclusion.....	176
Table 1. TG repeats polymorphisms and pathologies.....	16
Table 2. Nucleotide sequences enriched in TDP-43 targets.....	37
Table 3. Exons carrying UG/TG repeats at their 5'SS.....	119
Table 4. Oligonucleotides List.....	170

Supplementary Figures

[illegible]

☐ TDP-43 ☐ SC35
☐ SRp40 ☒ SRp55
☐ SF2/ASF

Figure 29. ESE and ISE sequences ensuring RXRG exon 7 constitutive inclusion. The *cis*-elements within the sequence representing *trans*-acting proteins-binding sites are represented with different colored lines on the top of the sequences.

☐ TDP-43 ☐ SC35
☐ SRp40 ☒ SRp55
☐ SF2/ASF

176